

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA**

**SAS VERSÃO 8.2
NA ANÁLISE DE EXPERIMENTOS:
CURSO BÁSICO**

N.º 62

Prof. Dr. João Riboldi

Prof^a. Dr^a. Dinara W. Xavier Fernandez

Karina Pretto

Michele Hartmann Feyh

Luana Beck

Bianca Pelicioli Riboldi

DEZEMBRO 2001

02 MAR. 2003

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
CADERNOS DE MATEMÁTICA E ESTATÍSTICA
SÉRIE A: TRABALHO DE PESQUISA

SAS VERSÃO 8.2 NA ANÁLISE DE EXPERIMENTOS: CURSO BÁSICO

JOÃO RIBOLDI
DINARA WESTPHALEN XAVIER FERNANDEZ
KARINA PRETTO
MICHELE HARTMANN FEYH
LUANA BECK
BIANCA PELICOLI RIBOLDI

SÉRIE B, Nº 62
PORTO ALEGRE, AGOSTO DE 2002

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
BIBLIOTECA SETORIAL DE MATEMÁTICA
SEÇÃO DE PERIÓDICOS

SUMÁRIO

1. O Sistema SAS	i
2. Analyst	1
3. Começando Com o Analyst	3
4. Trabalhando Com Dados	12
5. Transformando Dados	23
6. Combinando Tabelas	28
7. Trabalhando Com Colunas	34
8. Gráficos	40
9. Tarefas Estatísticas	64
9.1. Estatística Descritiva	64
9.2. Análise De Tabelas	84
9.3. Teste De Hipótese	91
9.3.2. Teste t para Média de uma Amostra	91
9.3.3. Teste t para Duas Amostras Pareadas	97
9.3.4. Teste t para a Diferença Entre Médias de Duas Amostras	102
9.4. Análise De Variância (ANOVA)	105
9.4.2. Análise de Variância em Classificação Simples (One-Way ANOVA)	107
9.4.3. Análise de Variância para Experimentos Fatoriais	112
9.5. Regressão	124
9.5.2. Regressão Linear Simples	124
9.5.3. Regressão Linear	129
10. Solução dos Exercícios	136

1. O SISTEMA SAS

A Versão 8.2 do Sistema SAS , caracteriza-se, especialmente pelo software de armazenamento de dados, que junta todas as ferramentas que você precisa para soluções analíticas, exploração de dados, visualização de negócios, desenvolvimento rápido de aplicações e muito mais. O SAS fornece um grupo de soluções prontas para usar e aplicações. Você pode acessar essas soluções usando o menu que se abre Soluções (Solutions). Selecione um dos tópicos abaixo para ir à informação desejada:

Análise (Analysis)

- 3D Visual Analysis
- Analyst
- Design of Experiments
- Geographic Information System :
- Guided Data Analysis
- Interactive Data Analysis
- Investment Analysis
- Market Research
- Project Management
- Quality Improvement
- Queuing Simulation
- Time Series Forecasting System
- Time Series Viewer

Desenvolvimento e Programação (Development and Programming)

- EIS/OLAP Application Builder
- Frame Builder
- Class Browser
- Source Control Manager
- Warehouse Administrator
- OLAP Server Administration

Repertório (Reporting)

- EIS/OLAP Report Gallery
- Design Report
- Graph-N-Go

Acessórios (Accessories)

- Graphic Test Pattern
- Locale Setup
- Registry Editor

ASSIST

Desktop

EIS/OLAP Application Builder

Usando o SAS Solutions:

Analyst: que será tratado no presente curso

Design of Experiments: solução de planejamento e análise de experimentos usando o ADX

Usando o Software SAS/GIS

O software SAS/GIS é um sistema de informações geográficas (GIS) que o permite exibir, analisar, e interagir com dados regionais tais como mapas e atribuir dados como número de população, densidade demográfica, ou contagem de tráfego.

Análise de Dados Guiado com o Software SAS/LAB

O software SAS/LAB fornece um ambiente integrado para executar análises de dados padrão, assim como regressão e análise de variação. O software aplica e interpreta resultados de análise, faz sugestões para análises futuras, e faz advertências sobre suposições violadas. A interface point-and-click fornece métodos convenientes para conversão de dados 'crus' para forma utilizável, produzindo resultados gráficos, e transformando resultados em relatórios.

O software SAS/LAB serve a vários profissionais:

- para o engenheiro ou cientista pesquisador que precisa analisar dados mas não é um estatístico;
- para os estatísticos consultores para apoio como profissionais;
- qualquer analista de dados que pode se beneficiar através de resultados gráficos e checagem automática de hipóteses.

O Software SAS/INSIGHT

Um Sistema iterativo de análise.

Muito embora os usuários busquem preferentemente as soluções prontas, não devemos esquecer que o Sistema SAS é composto de módulos, possuindo uma linguagem poderosa de programação, com uma coleção de programas chamados de **PROCEDURES(PROCS)**, disponíveis para a solução de problemas específicos

2. ANALYST

2.1 Caracterização

O Analyst é uma ferramenta de análise de dados que fornece um fácil acesso à análises estatísticas básicas. A aplicação é direcionada a estudantes e usuários novatos do SAS bem como para pesquisadores e analistas que estão familiarizados com o sistema SAS.

A interface inclui uma grande variedade de procedimentos analíticos e gráficos. Pode-se calcular estatísticas descritivas, executar testes de hipóteses, e ajustar modelos de análise de variância e de análise de regressão. Há também procedimentos para análise de sobrevivência, modelos mistos, análise de medidas repetidas e algumas técnicas multivariadas. O Analyst também fornece alguns procedimentos não coberto pelas PROC do SAS tais como alguns testes de hipóteses e determinação do tamanho de amostra e do cálculo do poder de testes estatísticos. Os gráficos são incluídos na maioria dos procedimentos analíticos, e você pode solicitar muitos tipos de gráficos diretamente.

Há muitos modos de entrar com os dados que você quer analisar. Você pode entrar com um conjunto de dados existente no SAS, ou abrindo os dados de uma base que não seja o SAS, ou ainda, digitando os valores você mesmo. Os dados são mostrados em uma tabela de dados em que as colunas correspondem às variáveis e as linhas correspondem às observações ou casos. Você pode editar elementos individuais na tabela de dados e criar novas variáveis. Você pode acessar aos procedimentos clicando no menu ou numa barra de ferramentas personalizada ou ainda através de um índice de termos frequentemente usados.

Os resultados das análises e gráficos são apresentados em janelas separadas, e todas estas saídas são controladas numa árvore do projeto, que pode conter os resultados de muitos conjuntos de dados. Os resultados podem ser apresentados como nós desta árvore, e você pode executar operações como desmanchar ou expandir ligações, deletar ligações, ou imprimir os resultados das ligações.

As tarefas são executadas com combinações de procedimentos SAS e passos de programação de dados. O código (CODE) atual submetido está disponível como uma ligação na árvore do projeto. O código pode ser usado pelos usuários que querem fazer uma auditoria de uma solução, ou interessados em aprender a programação SAS ou usar o procedimento de programação em seus próprios programas.

2.2 Novas Estruturas Do Analyst

A partir da versão 7 do SAS, o Analyst adicionou algumas novas vantagens incluindo novas análises estatísticas, novas possibilidades de gerenciamento de seus dados e de seus resultados.

Projetos

Seus resultados estão agora organizados em um projeto Analyst. Um projeto é uma coleção de resultados de análises executadas em um ou mais conjuntos de dados, por exemplo, os resultados de análises dentro de um estudo que envolve múltiplos conjuntos de dados. Um projeto é mostrado graficamente em uma árvore de projetos. Você pode criar, abrir, salvar, e deletar projetos do item **PROJETO** do menu **FILE**.

Exibir, Título e Configurações gráficas

Você pode agora selecionar títulos, configurações de exibição e configurações gráficas diretamente do menu **TOOLS**.

Abrindo e Salvando Arquivos Não-SAS

Você pode abrir e salvar arquivos em qualquer formato que não-SAS que seja suportado pelo Software SAS e sua plataforma. Por Exemplo, você pode salvar uma tabela de dados do SAS para a planilha EXCEL do Windows 95. A partir do menu **FILE**, selecione **OPEN** para abrir os dados não-SAS ou **SAVE AS** para salvar uma tabela de dados para um formato diferente.

Concatenando e Unindo Tabelas

Você pode unir tabelas pelas colunas ou concatená-las por linhas. Selecione **Combine Tables** do menu **DATA**, e selecione **Merge By Columns** ou **Concatenate By Rows**.

Transformando Dados

Você pode aplicar transformações assim como log e raiz quadrada das colunas para criar colunas calculadas, gravar valores de dados e amplitudes, converter o tipo de coluna para caracter ou numérica, e resumir colunas por grupo. Selecione os itens do menu **Data** para aplicar estas novas transformações em seus dados.

Trabalhando com Colunas

Você pode empilhar colunas para reordenar a forma de sua tabela de dados. Isto é particularmente usado quando você precisa reordenar dados de medidas repetidas antes de executar a análise. Você pode separar colunas para o produto de uma nova coluna não importa quando o valor da variável muda. Selecione **Stack Columns** ou **Split Columns** do menu **Dados**.

Submetendo o Código

Agora é mais fácil submeter o código (CODE) gerado na janela do Programa Editor abrindo a janela de código e selecionando **Copy to Program editor** do menu **Edit**.

Resultado HTML

Desde a Versão 7, você tem a possibilidade de criar arquivos HTML contendo seus resultados. A única tarefa que você não pode criar arquivos HTML são os **Hipotesis Test** do menu **Statistics**.

Documentos

Você pode escolher entre cinco diferentes estilos de tabelas de documentos que melhor refletem a proximidade entre seus dados. Selecione **Tables** do menu **Reports** para criar um documento arrumado em tabelas.

Gráficos

Você pode agora criar gráficos de setores e de barras horizontal e vertical.

Gerando Distribuições

Você pode gerar colunas de variáveis aleatórias de distribuições comuns, tal como Normal, Uniforme e Qui-quadrado. Selecione **Random Variates** do menu **Dados**.

Tarefas Estatísticas

Um número de novas tarefas estatísticas tem sido adicionadas ao menu **Statistics**:

- ANOVA: Medidas Repetidas
- ANOVA: Modelos Mistos
- Multivariada: Componentes Principais
- Multivariada: Correlação Canônica
- Análise de Sobrevivência: Tabelas de Vida
- Análise de Sobrevivência: Riscos Proporcionais

Além disso, na Versão 8, a ANOVA One-Way Não-paramétrica é agora uma tarefa separada.

3. Começando com o Analyst

3.1 Analyst:

Iniciamos a aplicação do Analyst selecionando **Analysis** no menu **Solutions** e posteriormente selecionando **Analyst**. Você também pode dar início ao Analyst de dentro do SAS escrevendo **Analyst** no comando *prompt*. Você pode exibir somente o Analyst quando executar um comando SAS por uso da opção *initcmd analyst*.

3.2 Menus:

Pode-se usar o Menu do Analyst para executar uma variedade de tarefas.

File Menu

Pelo **File** menu, você pode acessar e salvar arquivos, projetos, e conjunto de dados, e imprimir relatórios.

Edit Menu

Pelo **Edit** menu, você pode mudar os dados do modo *Browse* para o *Edit*, e adicionar, duplicar e delatar colunas e linhas.

View Menu

Pelo **View** menu, você pode mover e ocultar colunas, e modificar atributos da tabela de dados.

Tools Menu

Pelo **Tools** menu, pode colocar os títulos para seus resultados, gerar um conjunto de dados amostrais, especificar a visualização e preferências dos gráficos, e determinar uma nova SAS library.

Data Menu

Pelo **Data** menu, pode-se filtrar, ordenar, resumir, concatenar, juntar e aplicar análises em seus dados.

Reports Menu

Pelo **Reports** menu, você pode produzir detalhes e informar resumidamente.

Graphs Menu

Pelo **Graphs** menu, pode-se gerar gráficos.

Statistics Menu

Pelo **Statistics** menu, você pode escolher análises estatísticas, e usar o index para procurar por tarefas ou estatísticas.

Help Menu

Pelo **Help** menu, você pode exibir a ajuda para o Analyst e para o restante do programa SAS.

3.3 Barra de Ferramentas

A Barra de Ferramentas do Analyst contém ícones que você pode clicar para executar uma série de tarefas do Analyst. Por default, você pode executar essas ações pela Barra de Ferramentas:



Figura 3.1 Barra de Ferramentas

- Criar uma nova tabela
- Abrir um arquivo
- Salvar a tabela atual
- imprimir a tabela de dados
- visualizar a impressão
- classificar linhas em ordem ascendente
- escolher o local onde colocar o subconjunto de dados
- computar uma nova coluna
- criar um resumo
- criar um histograma
- criar um gráfico de dispersão bidimensional
- calcular estatísticas descritivas
- calcular correlações
- executar Análise de Variância em classificação simples (One-Way)
- executar uma regressão linear
- obter ajuda do Analyst.

3.3.1. Configurando a Barra de Ferramentas

Você pode adaptar a gosto pessoal a Barra de Ferramentas adicionando suas tarefas e ícones. Selecione **Customize** no Menu Ferramentas para exibir o diálogo da Ferramenta **Customize**.

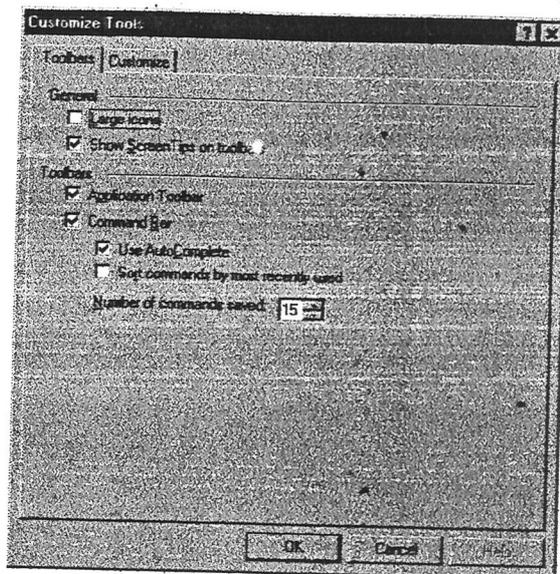


Figura. 3.2 Barra de Ferramentas

(Abaixo de Unix, selecione **Options** no Menu **Tools** (Ferramentas) e selecione **Edit Toolbox** para exibir o diálogo do **Tool Editor**)

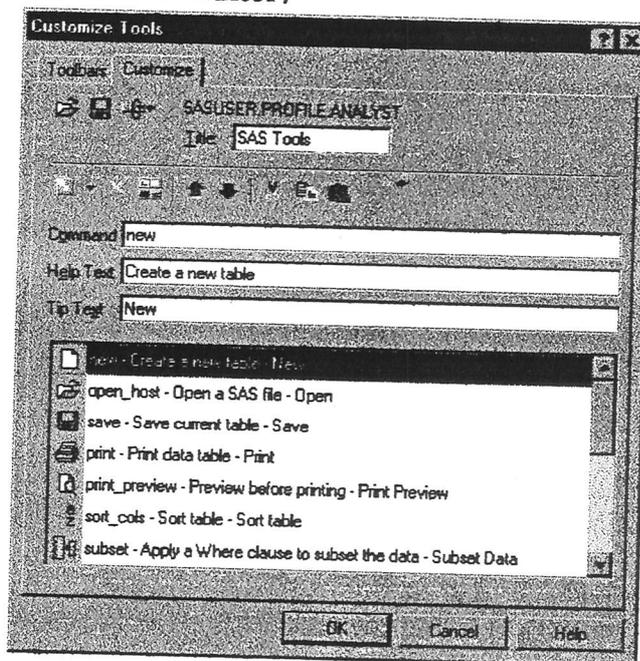


Figura 3.3 Personalizar Barra de Ferramentas

Comandos File

Tabela 3.1 Comandos File

Task	Command
New	NEW
Close	END
Open	OPEN HOST
Open By SAS Name	OPEN SAS
Open With New Query	QUERY WINDOW
Open With Existing Query	QUERY LIST
Save	SAVE
Save As	SAVEAS HOST
Save As By SAS Name	SAVEAS SAS
Projects	

New	NEW PROJECT
Open	OPEN PROJECT
Save	SAVE PROJECT
Save As	SAVE PROJECT AS
Delete	DELETE PROJECT
Print Preview	PRINT PREVIEW
Print Setup	PRINT SETUP
Print	PRINT

Comandos Edit

Tabela 3.2 Comandos Edit

Task	Command
Insert Columns	ADDCOLS
Add Rows	ADDRWS
Duplicate	DUPLICATE
Delete	DELETE
Rename	RENAME
Mode	
Browse	BROWSE MODE
Edit	EDIT MEMBER MODE
Shared Edit	EDIT RECORD MODE

Comandos View

Tabela 3.3 Comandos View

Task	Command
Columns	
Move	MOVE COLS
Hide	HIDE COLS
Unhide	UNHIDE COLS
Hold	HOLD COLS
Labels	SHOW LABELS
Table Attributes	TABATTRS

Comando Tools

Tabela 3.4 Comandos Tools

Task	Command
Titles	STITLES
Sample Data	SAMPLE DATA
Viewer Settings	PREFS
Graph Settings	GRAPH PREFS
New Library	LIB ASSIGN

Comandos Reports Table

Tabela 3.5 Comandos Tools

Task	Command
List Data	LIST DATA
Tables	TABLES

Comandos Graphs

Tabela 3.6 Comandos Tools

Task	Command
Bar Chart	
Horizontal	HBAR
Vertical	VBAR
Pie Chart	PIECHART
Histogram	HIST

Box Plot	BOX
Probability Plot	NORMPLOT
Scatter Plot	
Two-Dimensional	SCAT2D
Three-Dimensional	SCAT3D
Contour Plot	CONTOUR
Surface Plot	SURFACE

Comando Data

Tabela 3.7 Comandos Data

Task	Command
Filter	
None	SUBSET CLEAR
Subset Data	SUBSET
Sort	SORT COLS
Transform	
Compute	COMPUTED COLUMN
Rank	RANK
Standardize	STANDARDIZE
Recode Values	RECODE VALUES
Recode Ranges	RECODE RANGES
Convert Type	CONVERT TYPE
Log(Y)	TRN LOG
Sqrt(Y)	TRN SORT
1/Y	TRN RECIP
Y*Y	TRN SQUARE
Exp(Y)	TRN EXP
Random Variates	
Normal	RV NORMAL
Uniform	RV UNI
Binomial	RV BIN
Chi-Square	RV CHI
Poisson	RV POIS
Beta	RV BETA
Exponential	RV EXP
Gamma	RV GAMMA
Geometric	RV GEOM
Extreme Value	RV EXTREME
Summarize By Group	SUM BY GROUP
Combine Tables	
Merge By Columns	MERGE
Concatenate By Rows	CONCATENATE
Stack Columns	STACK
Split Columns	SPLIT
Transpose	TRANSPOSE
Random Sample	RANSAMP
Column Properties	COLATTRS

Comandos Statistics

Tabela 3.8 Comandos Statistics

Task	Command
Descriptive	
Summary Statistics	SUMMARY
Distributions	DISTRIB
Correlations	CORR
Frequency Counts	COUNTS

Table Analysis	TABLANAL
Hypothesis Tests	
One-Sample Z-test for a Mean	HT1Z
One-Sample t-test for a Mean	HT1T
One-Sample Test for a Proportion	HT1P
One-Sample Test for a Variance	HT1V
Two-Sample t-test for Means	HT2T
Two-Sample Paired t-test for Means	HT2PT
Two-Sample Test for Proportions	HT2P
Two-Sample Test for Variances	HT2V
ANOVA	
One-Way ANOVA	ONEANOVA
Nonparametric One-Way ANOVA	NONPARAM
Factorial ANOVA	FACANOVA
Linear Models	LINMOD
Repeated Measures	RMANOVA
Mixed Models	MIXED
Regression	
Simple	SIMPREGR
Linear	LINREGR
Logistic	LOGREGR
Multivariate	
Principal Components	PRINCOMP
Canonical Correlation	CANCORR
Survival	
Life Tables	LIFETEST
Proportional Hazards	PHREG
Sample Size	
One-Sample t-test	SSPMEAN1T
One-Sample Confidence Interval	SSPMEAN1CI
One-Sample Equivalence	SSPMEAN1E
Paired t-test	SSPMEANPT
Paired Confidence Interval	SSPMEANPCI
Paired Equivalence	SSPMEANPE
Two-Sample t-test	SSPMEAN2T
Two-Sample Confidence Interval	SSPMEAN2CI
Two-Sample Equivalence	SSPMEAN2E
One-Way ANOVA	SSPMEANIA
Index	INDEX

Comando Help

Tabela 3.9 Comandos Help

Task	Command
Using This Window	window help

3.4 Índice do diálogo de Tarefas

Use o index para acessar tarefas estatísticas e gráficos através de termos geralmente usados.

Selecione **Index** no Menu **Statistics** para exibir o Índice do Diálogo Tarefas. Clique na palavra para abrir uma tarefa, ou escreva a tarefa no campo para encontra-la na lista. Por exemplo, se você quiser usar o teste de Brown-Forsythe, clique **Brown-Forsythe test of equal variances**, clique em **OK**, e a tarefa ANOVA estará aberta. O Teste de Brown-Forsythe estará disponível no Diálogo.

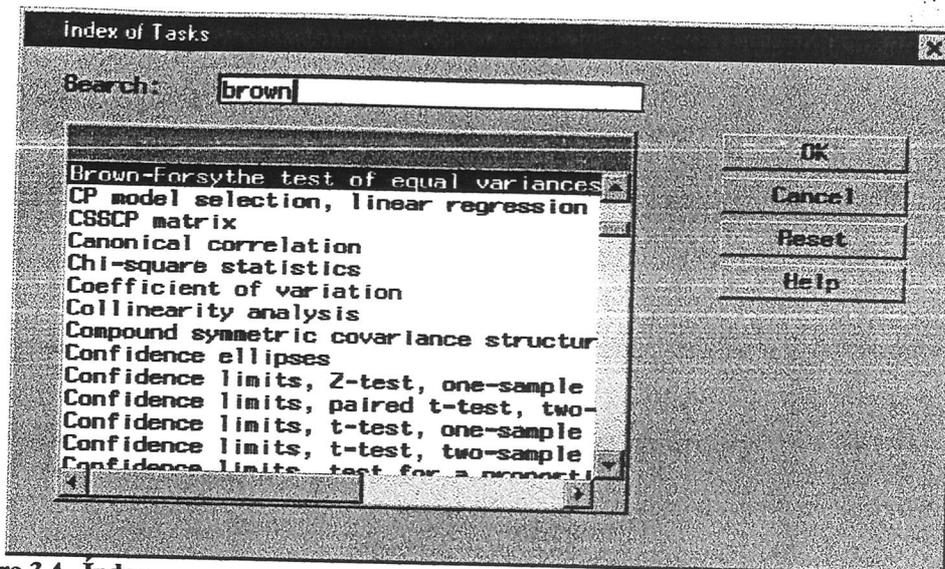


Figura 3.4 Index

3.5 Projetos

Um projeto do Analyst é uma coleção de resultados de análises executadas de um ou mais grupos de dados, por exemplo, os resultados das análises dos dados de um estudo que englobe múltiplos conjunto de dados.

Um projeto é exibido como uma árvore de projetos que contenham pastas de diferentes tabelas de dados, relatórios, códigos e outros resultados que são associados com o projeto. Os resultados são apresentados como uma ligação (nó) desta árvore.

Você pode abrir e fechar uma pasta clicando no sinal de mais (+) ou de menos (-) ou selecionando a pasta e selecionando **Expand** (expandir) ou **Collapse** (ocultação de subdiretórios) no Menu desdobrável.

Você pode ver uma ligação dentro da pasta por um clique duplo na ligação, ou selecionando **View** (ver) no Menu desdobrável. Você pode ver a mesma ligação em duas ou mais janelas selecionando a ligação e selecionando **View in new Window** (visualizar em nova janela) no Menu desdobrável.

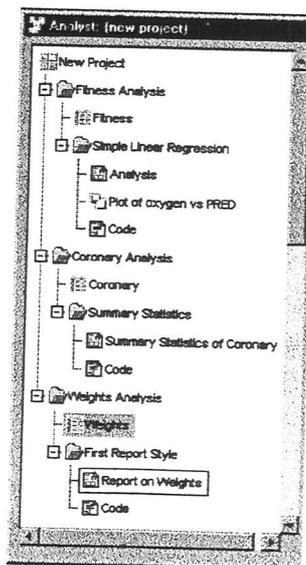


Figura 3.5 Árvore de Projeto

Pelo Menu desdobrável, você pode também deletar, imprimir, salvar e, exceto em casos de tabelas de dados, renomear a ligação.

Se a ligação é uma tabela, você pode visualizar a tabela em uma nova janela selecionando **View** (ver) no Menu desdobrável, ou você pode abrir a tabela para análise selecionando **Open** (abrir) no Menu desdobrável. Também, você pode selecionar **Interactive Analysis** (Análises Interativas) no Menu desdobrável para executar o comando SAS/INSIGHT® software (se estiver instalado) para executar a análise exploratória interativa.

Para criar, abrir, salvar ou deletar o Projeto de Análise, selecione **Projects** no menu **File**.

3.6 Botões

O diálogo Analyst possui botões simples que executam as mesmas ações em cada diálogo.

Ok Button

Clique no botão **OK** para aplicar as mudanças especificadas e para fechar o diálogo.

Cancel Button

Clique no botão **Cancel** para cancelar suas mudanças e sair do diálogo.

Reset Button

Clique em **Reset** para restaurar os valores no diálogo preservando os valores defaults ou para as configurações que você necessita salvar com o botão **Save Options** (Salvar opções). O botão **Reset** geralmente remove as variáveis no diálogo para determinada função.

Save Options Button

Clique em **Save Options** para salvar qualquer configuração que você tiver especificado. Estas configurações tornam-se seus defaults, e voltarão as configurações determinadas quando você clicar o botão **Reset**. Variáveis tarefas não são salvas.

3.7 Visualizando e salvando códigos

Para alguma tarefa estatística ou gráfico que você quiser aplicar aos seus dados no Analyst, você pode visualizar e salvar o código (programação). Este código original pode ser copiado dentro do editor de programa e editado e submetido posteriormente.

Selecionando **Viewer Settings** (visualizar ambiente) no Menu **Tools** (Ferramentas), e clicar na etiqueta **Output** (resultado). Selecionando **Provide source code** (providenciar código de origem) para incluir um item no código de origem em seu projeto árvore toda vez que você aplicar uma tarefa em seus dados.

Para resubmeter um código na janela de edição de programas, dê um clique duplo no item do código no projeto árvore para ver o código na janela Code. Selecione **Copy to Program Editor** (copiar para a janela do Programa Editor) no Menu **Edit**. Os dados precisam estar ramificados na ordem para você submeter o código na tabela de dados corrente.

3.8 Visualizar resultados em múltiplas janelas

Você pode executar análises em uma ou mais tabelas de dados e ver os resultados simultaneamente em múltiplas janelas.

Para visualizar um resultado em uma única janela, dê um clique duplo no resultado na árvore do projeto ou selecionando o resultado e selecionando **View** (ver) no Menu desdobrável.

Se você aplicar outras análises para a tabela de dados, os resultados das análises são armazenados em pastas separadas, e você pode visualizar os resultados da segunda análise numa janela separada.

Você também pode abrir uma nova janela ligado ao mesmo resultado selecionando o resultado e selecionando **View in new Window** (ver em nova janela) no Menu desdobrável.

3.9 Salvando opção de tarefas

Você pode salvar alguma opção que está associada com a tarefa clicando no botão **Save Options** (salvar opção) no diálogo de tarefas. Por exemplo, você pode salvar as opções que são associadas com a tarefa Gráfico de Barras clicando no botão **Save Options** no **Bar Chart dialog** (diálogo do gráfico de barras). Esta opção tornará seu default e serão aplicadas quando você clicar no botão **Reset** (retornar). Estas opções também podem ser salvas no entre sessões.

Opções que são associadas com dados, tais como **Group By variables** (grupo por variáveis) não podem ser salvas com as opções de tarefas, e não continuam entre sessões.

3.10 Salvando textos e resultados gráficos

Você pode salvar códigos, análises e resultados gráficos como arquivos. Para salvar códigos ou resultados, selecione a ligação na árvore de projetos e selecione **Save as** (salvar como) no Menu desdobrável. Especificar um nome para o arquivo e selecione tipo para o arquivo. Para resultados gráficos, você pode salvar a figura em formatos que incluem GIF, JPEG, and Postscript. Clique em **Save** (salvar) para salvar o código ou resultado.

Você também pode salvar códigos ou resultados dando um clique duplo na ligação para abrir, e selecionar **Save As** (salvar como) no menu **File** (arquivo).

3.11 Personalizando sua Sessão

Você pode personalizar sua sessão Analyst no Menu **Tools** (ferramentas) selecionando **Viewer Settings** (visualizar ambiente) para observar séries de preferências e **Graph Settings** (ambiente gráfico) para observar as preferências gráficas. Opções gerais são canceladas por qualquer ambientação que você especificar na tarefa. Essas opções também são canceladas por opções que são salvas no botão **Save Options** (salvar opções) nas tarefas. Selecione **Titles** (título) no Menu **Tools** ou clique no botão **Titles** dentro da tarefa para especificar o título que aparecer no output. Títulos gerais são salvos através da sessão Analyst.

4. Trabalhando Com Dados

4.1 Abrindo Dados

A tabela de dados do Analyst mostra os dados do conjunto de dados do SAS, dados observados, e outras origens, semelhante à planilha eletrônica do Excel e arquivos portáteis do SPSS. Você pode também entrar com os dados na tabela de dados.

Você gera uma nova tabela de dados, abre o conjunto de dados do SAS e observa, e abre arquivos que não são do SAS do menu File.

New

Selecione **New** do menu **File** para gerar uma nova (vazia) tabela de dados.

Exercício1.

Criando um banco de dados.

Para entrar com os dados no Analyst, você pode digitá-los diretamente na planilha eletrônica.

1. Digite os seguintes dados na janela do Analyst:

20	40	160
23	50	165
25	60	168
30	55	170

2. A primeira coluna indica a Idade, o Peso e a terceira a Altura.
3. Salve o conjunto de dados no diretório de sua preferência, clicando no menu **File** e depois em **Save**.
4. Escolha um nome para o seu arquivo (ex1)
5. Para nomear as colunas, clique com o botão direito do mouse sobre o cabeçalho da coluna e então clique em **Properties**. Digite os nomes das colunas no campo **Name**.

Open

Selecione **Open** do menu **File** para abrir arquivos e conjuntos de dados do SAS e mostrar dados do seu diretório ou pasta de sistema operacional. Você pode trazer qualquer formato de arquivo que o SAS suporta na tabela de dados, incluindo conjunto de dados do SAS, Excel, Lotus e arquivos portáteis do SPSS.

Utilize o filtro **Files of type** no diálogo **Open** para selecionar o tipo de arquivo que se deseja. Selecione um arquivo, e clique em **Open** para trazer o conteúdo do arquivo para a tabela de dados. Arquivos que não são do SAS abertos no Analyst são convertidos em conjunto de dados do SAS. Os arquivos de origem não são alterados.

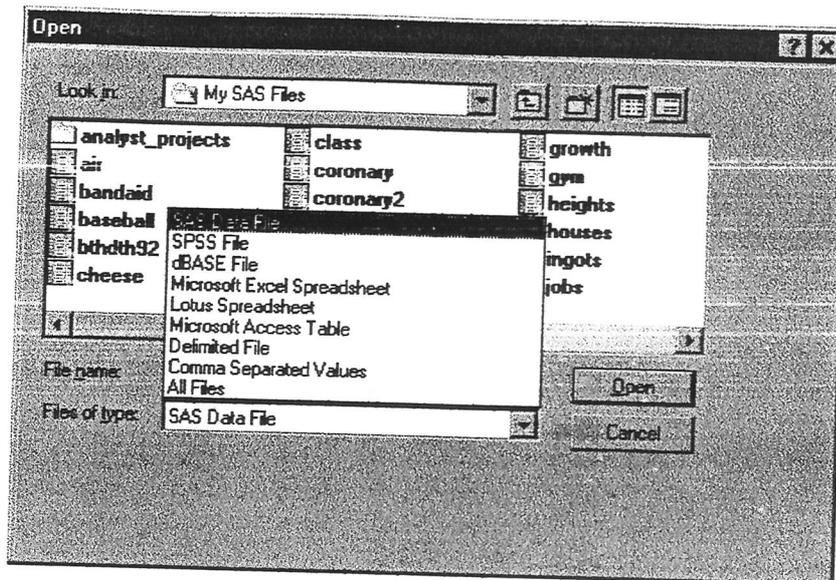


Figura 4.1 Diálogo Open

Exercício2.

Abrindo um arquivo já existente com extensão .SAS

1. Clique em **Open** do menu **File**.
2. Vá até o diretório onde o seu arquivo está salvo.(C:\ user\curso_sas)
3. Clique no arquivo **Populacao.SAS**.
4. Deixe selecionada a opção que exhibe o nome das variáveis na primeira linha.
5. Clique em **OK**.
6. Salve o arquivo no seu diretório de trabalho.

Você pode abrir planilhas eletrônicas de vários formatos do Excel, incluindo Excel 4, Excel 5, Excel 7 e Excel 97. Para trazer a planilha do Excel para dentro da tabela de dados, você deve se assegurar que as linhas correspondem às observações e que as colunas correspondem às variáveis. Você pode incluir o nome das variáveis na primeira linha da planilha eletrônica.

Para abrir uma planilha eletrônica do Excel no Analyst, selecione **File Open** e clique na seta **Files of Type** e selecione **Microsoft Excel Spreadsheet** da lista. Depois, vá até o diretório onde a planilha desejada se encontra e clique em **OK**.

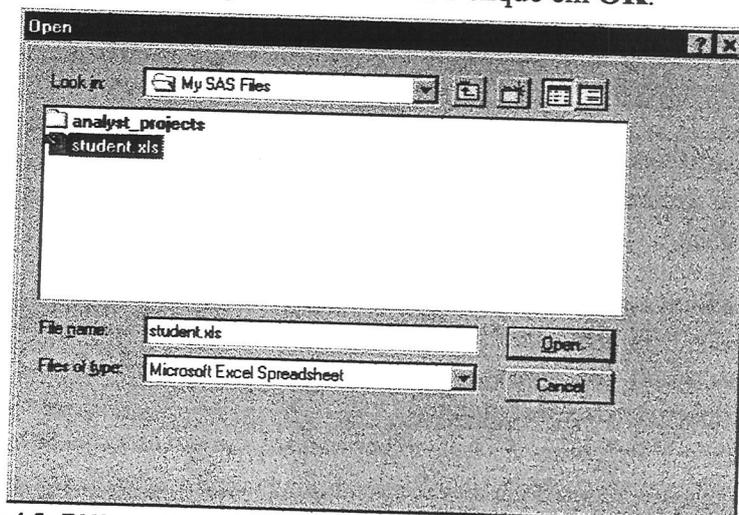


Figura 4.2. Diálogo Abrir com uma planilha eletrônica do Excel selecionada

Clique na seta **Worksheet/Range** e selecione a planilha de trabalho ou a linha de trabalho. Se a primeira linha da planilha não contém o nome das variáveis, desmarque **Column names in first row**. Clique em **OK** para abrir a planilha para dentro da tabela de dados.

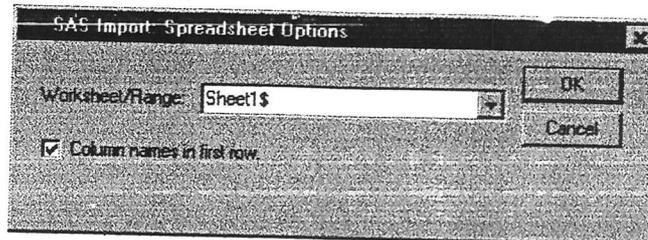


Figura 4.3 Importação SAS : Janela de Opções da Planilha Eletrônica

	name	sex	age	height	weight
1	Alice	F	13	56.5	84
2	Becka	F	13	65.3	98
3	Gail	F	14	64.3	90
4	Karen	F	12	56.3	77
5	Kathy	F	12	59.8	84.5
6	Mary	F	15	66.5	112
7	Sandy	F	11	51.3	50.5
8	Sharon	F	15	62.5	112.5
9	Tammy	F	14	62.8	102.5
10	Alfred	M	14	69	112.5
11	Duke	M	14	63.5	102.5
12	Guido	M	15	67	133
13	James	M	12	57.3	83
14	Jeffrey	M	13	62.5	84
15	John	M	12	59	99.5
16	Philip	M	16	72	150
17	Robert	M	12	64.8	128
18	Thomas	M	11	57.5	85
19	William	M	15	66.5	112

Figura 4.4 Tabela de Dados com Dados de uma Planilha Eletrônica do Microsoft Excel

Exercício3.

Abrindo uma planilha eletrônica do Microsoft Excel

1. Clique em **Open** do menu **File**.
2. Vá até o diretório onde o seu arquivo está salvo.(C:\ user\curso_sas)
3. No campo **Arquivos do Tipo** selecione **Microsoft Excel Spreadsheet**.
4. Clique no arquivo **Exercicio3**.
5. Deixe selecionada a opção que exibe o nome das variáveis na primeira linha.
6. Clique em **OK**.
7. Salve o arquivo no seu diretório de trabalho com o nome de **peso.sas**.

O Analyst também fornece acesso a dados de arquivos de texto. Arquivos de texto geralmente contém o nome das variáveis na primeira linha e as observações nas linhas seguintes. Cada nome de variável e valor dos dados está separado por uma vírgula, ponto e vírgula tab, ou outro delimitador.

```

name,sex,age,height,weight
Alice,F,13,56.5,84
Becka,F,13,65.3,98
Gail,F,14,64.3,90
Karen,F,12,56.3,77
Kathy,F,12,59.8,84.5
Mary,F,15,66.5,112
Sandy,F,11,51.3,50.5
Sharon,F,15,62.5,112.5
Tammy,F,14,62.8,102.5
Alfred,M,14,69,112.5
Duke,M,14,63.5,102.5
Guido,M,15,67,133
James,M,12,57.3,83
Jeffrey,M,13,62.5,84
John,M,12,59,99.5
Philip,M,16,72,150
Robert,M,12,64.8,128
Thomas,M,11,57.5,85
William,M,15,66.5,112

```

Figura4.5 Arquivos delimitados com Valores separados com vírgula.

Para abrir dados provenientes de arquivos de texto, selecione **File Open**, clique na seta **Files of Type** e depois selecione **Delimited File** da lista. Vá até o diretório em que o arquivo de texto está armazenado, selecione-o e clique em **Open**. Uma vez que você selecionou, você deve especificar como este arquivo está delimitado na janela **SAS Import: Delimited File Options**.

Selecione o tipo de delimitador na caixa. Se o arquivo de texto não está separado por espaço ou tab, você pode especificar um delimitador personalizado como vírgula, dois pontos ou ponto e vírgula selecionando **Character** e escrevendo o delimitador no campo. Especifique o número de linhas e o que contém a primeira linha de dados. O padrão do número da linha é 2 que é apropriado para arquivos que contêm o nome das variáveis na primeira linha. Selecione **Get variable names from first row** se a primeira linha do arquivo contém o nome das variáveis. Clique em **OK** para abrir os dados dentro da tabela de dados.

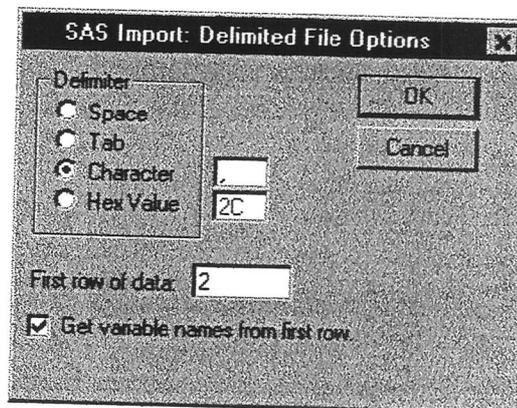


Figura 4.6 Janela Opções de Arquivos Delimitados

Exercício4.

Abrindo arquivos delimitados

1. Clique em **Open** do menu **File**.
2. Vá até o diretório onde o seu arquivo está salvo.(C:\user\curso_sas)
3. No campo **Arquivos do Tipo** selecione **Delimited Files**.
4. Clique no arquivo **Exercício4**.
5. Escolha o delimitador "character" e digite uma vírgula no campo indicado.
6. Clique em **OK**.
7. Salve o arquivo no seu diretório de trabalho.

Open By SAS Name

Selecione **Open By SAS Name** do menu **File** para abrir conjuntos de dados do SAS ou dados observados. Selecione a biblioteca do SAS e selecione um membro. Clique em **OK** para trazer o conteúdo do conjunto de dados do SAS ou os dados observados para a tabela de dados.

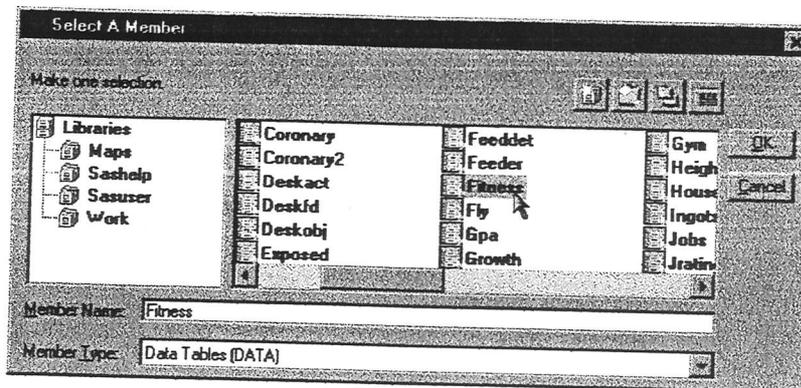


Figura 4.7 Diálogo Selecionando um Membro

Open With New Query

Você pode utilizar a janela **Query** para reduzir o número de variáveis que você carrega na tabela de dados. Você pode utilizar também para trazer mais de um conjunto de dados para a tabela de dados, assim como escrever consulta SQL para filtrar os dados. Selecione **Open With New Query** do menu **File** para abrir uma tabela do SAS por aplicação de uma nova consulta.

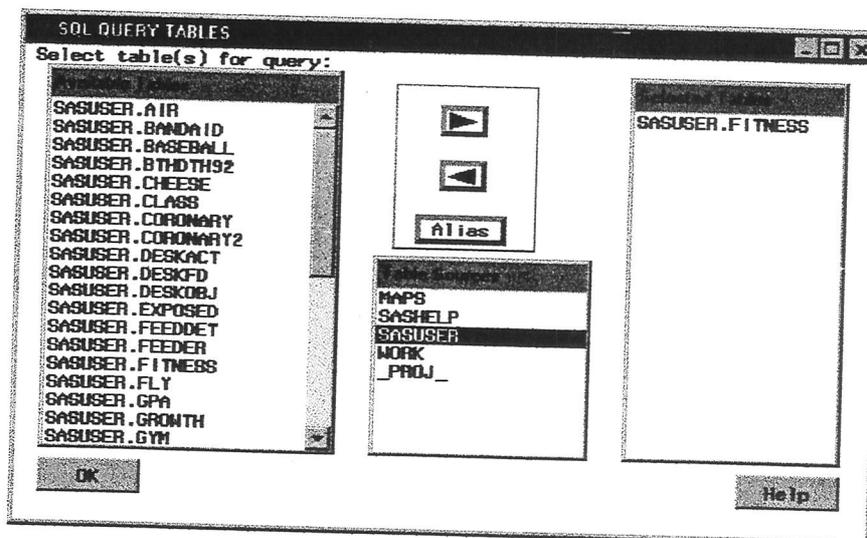


Figura 4.8 Janela da Tabela SQL QUERY

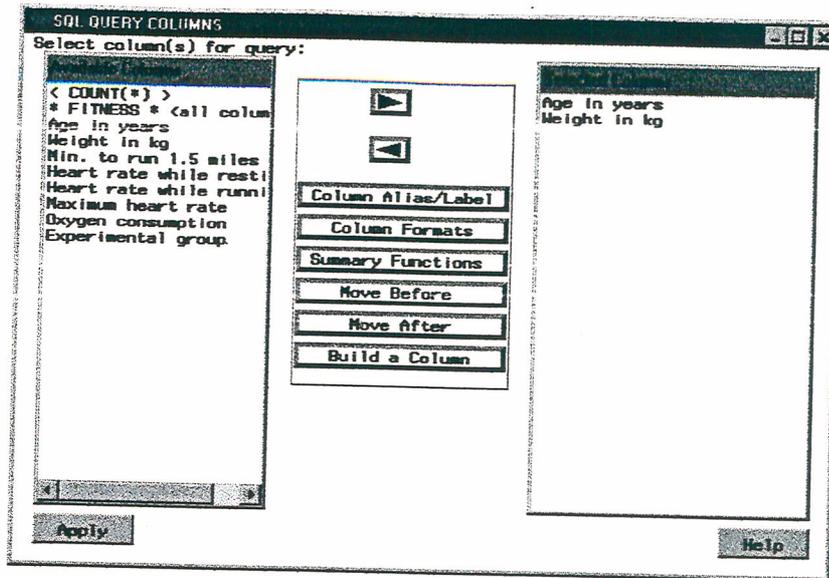


Figura 4.9 Janela das Colunas da SQL QUERY

Open With Existing Query

Selecione **Open With Existing Query** do menu **File** para abrir uma tabela do SAS por aplicação de uma consulta existente.

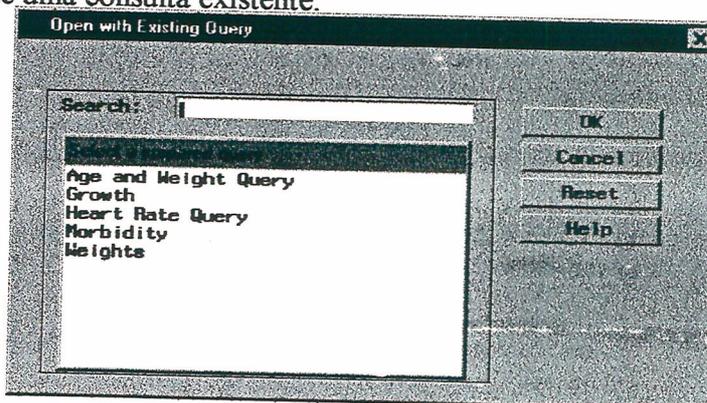


Figura 4.10 Open with Existing Query Window

4.2 A Tabela de Dados

Quando você abre um arquivo ou membro da biblioteca SAS no Analyst, os dados são trazidos para uma tabela de dados onde você pode observar e editar os dados e rearrumar as colunas e linhas. Se a tabela de dados é aberta nos modos procurar (**Browse**), editar(**Edit**) ou editar dividindo(**Shared Edit**) é determinada no diálogo **Viewer Settings** (ambiente de observação).

Browse Mode

Selecione **Mode** do menu **Edit**, e selecione **Browse** para prevenir mudanças para a tabela enquanto você está observando.

Edit Mode

Selecione **Mode** do menu **Edit**, e selecione **Edit** para mudar a tabela. Enquanto você está no modo **Edit**, ninguém mais está hábil para mudar a tabela.

Shared Edit Mode

Selecione **Mode** do menu **Edit**, e selecione **Shared Edit** para permitir que mais de uma pessoa faça mudanças simultaneamente na tabela. Os registros que você está editando são fechados enquanto você está editando, mas outros usuários podem mudar outros registros na tabela.

Typing Text

Quando você está nos modos **Edit** ou **Shared**, você pode mudar a tabela de dados selecionando uma célula e classificando-a na tabela.

Inserting Columns

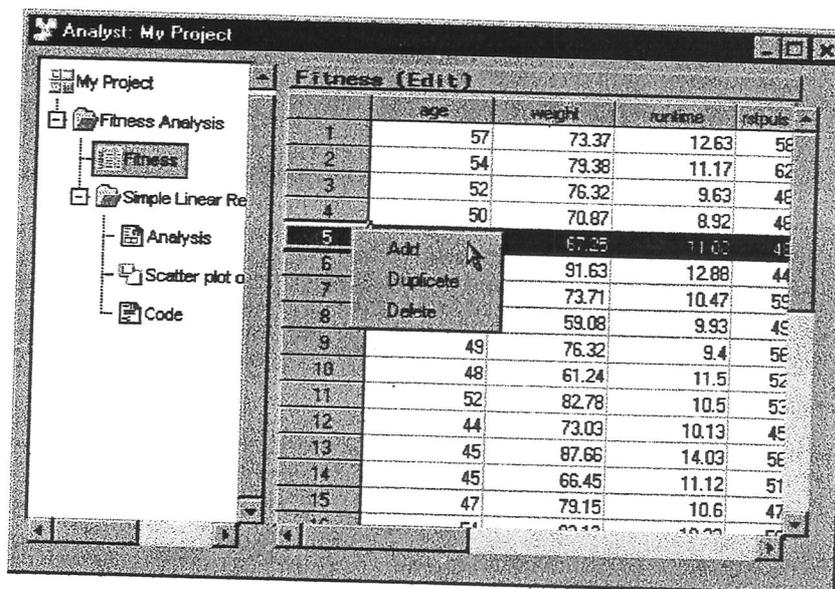
Para inserir uma coluna ao final da tabela, selecione **Inserting Columns** do menu **Edit**, ou selecione uma coluna e selecione **Inserting Columns** do menu . Se você selecionar uma coluna, uma nova coluna é adicionada antes da coluna selecionada.

Adding Rows

Para adicionar uma linha ao final da tabela, selecione **Add Rows** do menu **Edit**.

Deleting Rows

Para deletar uma linha, destaque a linha, e selecione **Delete** do menu **Edit**.



The screenshot shows the Minitab 'Fitness (Edit)' window. The data table is as follows:

	age	weight	runtime	stepuls
1	57	73.37	12.63	58
2	54	79.38	11.17	62
3	52	76.32	9.63	48
4	50	70.87	8.92	48
5		67.25	11.00	48
6		91.63	12.88	44
7		73.71	10.47	58
8		59.08	9.93	48
9	49	76.32	9.4	58
10	48	61.24	11.5	52
11	52	82.78	10.5	53
12	44	73.03	10.13	48
13	45	87.66	14.03	58
14	45	66.45	11.12	51
15	47	79.15	10.6	47

Figura 4.11 Menu Desdobrável

Duplicating Columns

Para duplicar uma coluna, selecione a coluna e selecione **Duplicate** do menu. A nova coluna é adicionada antes da coluna selecionada.

Deleting Columns

Para deletar uma coluna, selecione a coluna e selecione **Delete** do menu.

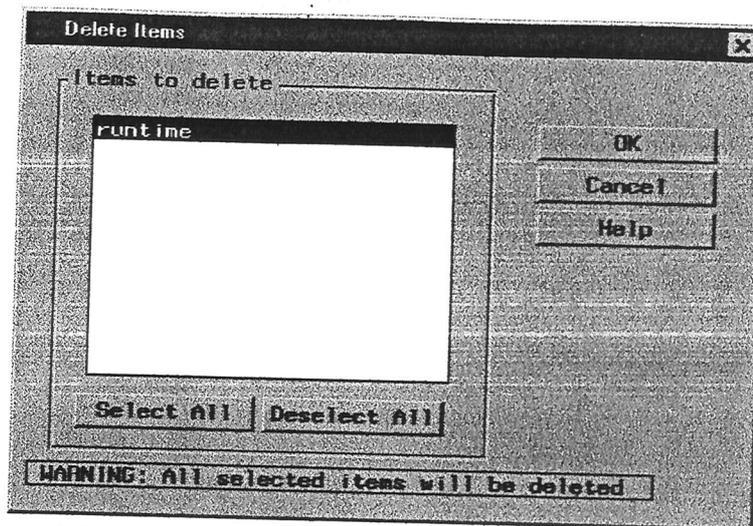


Figura 4.12 Diálogo Deletar Ítens

Moving Columns

Para mover uma coluna, selecione uma coluna e selecione **Move** do menu para mostrar a “Move Columns Dialog”.

Sorting Rows

Para ordenar linhas, selecione uma coluna e selecione **Sort** do menu para mostrar a “Sort Dialog”.

Hiding Columns

Para ocultar uma coluna, selecione a coluna e selecione **Hide** do menu para mostrar a “Hide Columns Dialog”.

Holding Columns

Para segurar algumas colunas enquanto ordena outras, selecione a coluna ou colunas que você deseja segurar e selecione **Hold** do menu para mostrar o “Hold Columns Dialog”.

Viewing Variable Names and Column Labels

Você pode optar entre exibir os nomes das variáveis e rótulos das colunas selecionando uma coluna e selecionando **Labels** do menu.

4.3 Salvando dados

Você pode salvar dados na tabela de dados para qualquer formato que o SAS suporte, incluindo conjuntos de dados do SAS, e arquivos do Excel e Lotus.

Salvando dados

Selecione **Save** do menu **File** para salvar mudanças feitas para o conjunto de dados atual.

Selecione **Save As** do menu **File** para salvar a tabela atual para um formato que não seja o SAS ou para salvar a tabela atual em outro conjunto de dados do SAS.

Utilize o filtro **Save as type** no diálogo **Save As** para selecionar o tipo de arquivo que você quer salvar. Clique em **Save** para salvar os conteúdos da tabela de dados para o arquivo.

Saving Data to a SAS Library

Selecione **Save As By SAS Name** do menu **File** para salvar a tabela atual como um conjunto de dados do SAS.

Selecione uma biblioteca. Selecione um conjunto de dados existente da lista de membros ou um tipo de nome de membro para um novo conjunto de dados no campo **Member Name**. Clique em **Save** para salvar o conjunto de dados.

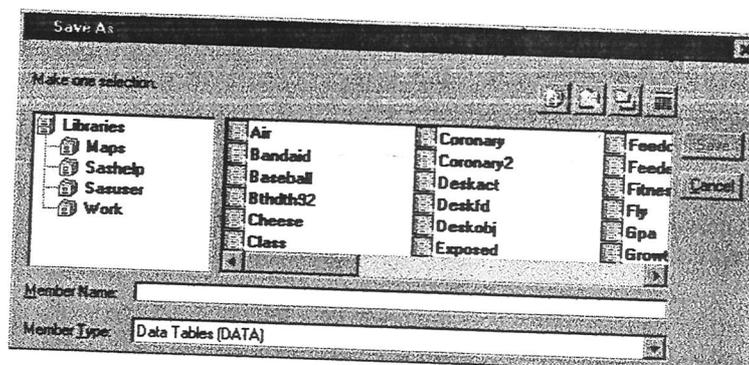


Figura 4.13 Diálogo Salvar Como

Reserved Names

Os nomes seguintes são reservados pelo Analyst e não devem ser utilizados para se referir à tabelas.

O `_proj_` libref aponta para a biblioteca do projeto atual onde os arquivos do projeto são armazenados.

Um `_tmp_` libref é determinado pelo Analyst como necessário. `_tmp_` é utilizado também como parte inicial de nomes para conjunto de dados temporários utilizados pelo Analyst, por exemplo, `_tmp_0439`.

4.4 Filtrando dados

Você pode ver um subconjunto de seus dados selecionando **Filter** do menu **Data**, e selecionando **Subset Data** para aplicar uma cláusula quando para sua tabela de dados. Selecione **None** se você não deseja o subconjunto dos seus dados, ou se você quer remover um sub-sistema existente. **None** é a ausência. Se você executa uma análise neste momento, a análise é executada no sub-sistema. Se você seleciona **Save** do menu **File**, os dados de toda a tabela são salvos. Se você seleciona **Save As** do menu **File**, somente o subconjunto é salvo. Se você modifica a célula para um valor fora dos critérios do sub-sistema, a linha é retirada da exibição da tabela. Se você, então, executa uma tarefa que utiliza o subconjunto, o subconjunto que é executado não incluindo a observação modificada.

4.5 Ordenando Dados

Selecione **Sort** do menu **Data** para ordenar as linhas na tabela dos dados. Selecione colunas da lista de candidatos e clique o botão **Sort by** para especificar as colunas a serem ordenadas.

Utilize as setas para cima e para baixo próximo à lista **Sort by** para especificar a ordem de seleção da coluna desejada. Selecione uma coluna e clique o botão **Ascend/Descend** para ordenar as linhas na coluna em ordem alfabética ascendente ou decendente. As linhas são ordenadas na ordem ascendente por definição.

Exercício 5.1:

Neste exercício, você tem um banco de dados que contém o código, o Estado e o número de cabeças de gado por região. Ordene o banco de dados em ordem alfabética do Estado.

1. Abra o banco de dados **rebanho_estado.sas** (pode ser salvo como **.SAS** ou como **.XLS** para exercitar mais uma vez como se abre arquivos com outro formato) salvo no seu diretório;
2. Troque o modo de exibição para **Edit** selecionando **Mode** no menu **Edit**.
3. A próxima etapa é selecionar o **Sort** do menu **Data**.
4. Selecione a variável **Estado** e depois clique no botão **Sort By**.
5. Clique em **OK** para executar a tarefa.

Exercício 5.2:

Ordene agora seu banco de dados descendente do número de cabeças de gado do estado.

6. A próxima etapa é selecionar o **Sort** do menu **Data**.
7. Selecione a variável **Cabeças** e depois clique no botão **Sort By**.
8. Selecione esta variável clicando uma só vez sobre ela e então clique no botão **Ascend/Descend**. Note que a letra que antecede a variável muda de **"(A)"** para **"(D)"**.
9. Clique em **OK** para executar a tarefa.

4.6 Calculando Distribuições

Você pode criar variáveis aleatórias baseadas em distribuições selecionando **Random Variates** do menu **Data**.

Você pode selecionar uma destas distribuições:

- ◆ Normal
- ◆ Uniforme
- ◆ Binomial
- ◆ Qui- quadrado
- ◆ Poisson
- ◆ Beta
- ◆ Exponencial
- ◆ Gama
- ◆ Geometrica
- ◆ Valor Extremo

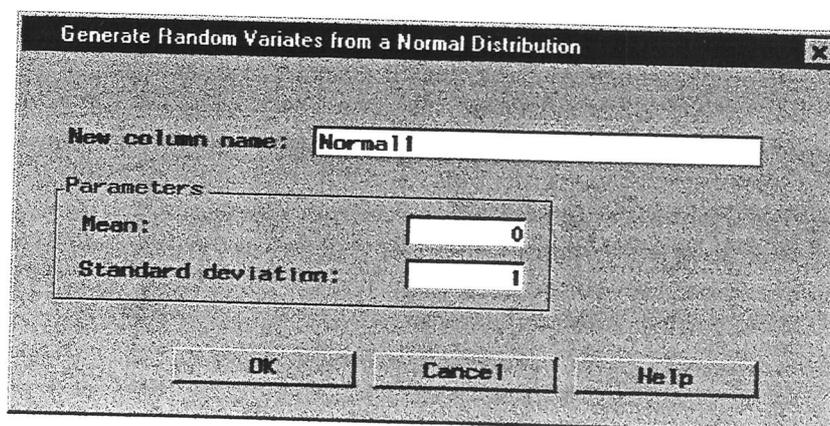


Figura 4.14 Diálogo Gerando uma Variável Aleatória com distribuição Normal

Especifique o parâmetro ou parâmetros da distribuição para o tipo que você selecionou e clique **OK**. Uma nova coluna que contenha as variáveis aleatórias é criada.

4.7 Resumir pelo Grupo de Diálogo

Selecione **Summarize by Group** do menu **Data** para mostrar o resumo por Grupo de diálogo. O Resumo por Grupo de diálogo produz uma saída do conjunto de dados que resume o conjunto de dados original. A saída do conjunto de dados contém uma linha para cada um dos níveis de um agrupamento de variáveis que você especifique. Se há muitos agrupamentos de variáveis, o resumo do conjunto de dados contém uma linha para cada combinação de níveis de agrupamentos de variáveis. As colunas no novo conjunto de dados são níveis de agrupamento de variáveis e as estatísticas selecionadas computadas para o grupo de observações definido por essa combinação. Se o agrupamento de variáveis não é selecionado, estatísticas são computadas para todas as observações.

Lista de Candidatos

A lista de candidatos mostra todas as variáveis na tabela de dados. Você pode selecionar variáveis desta lista para resumir ou utilizar como agrupamento de variáveis. Note que selecionando variáveis não contíguas da lista de candidatos, somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muito locais você pode então pressionar **Ctrl** enquanto clica nas variáveis.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto da lista de candidatos, a qual lista todas as variáveis no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista clicando duas vezes nelas.

Botão Summarize

Selecione variáveis da lista de candidatos e clique no botão **Summarize** para resumir as variáveis. Você deve selecionar pelo menos uma variável.

Botão Grupo

Selecione a variável da lista de candidatos e clique no botão **Group** para adicionar as variáveis à serem utilizadas como agrupamento de variáveis.

Estatísticas

Selecione uma ou mais estatísticas para adicionar ao conjunto de dados resumido. Você deve selecionar pelo menos uma estatística.

4.8 Diálogo Amostras Aleatórias

Selecione **Random Sample** do menu **Data** para criar uma amostra aleatória dos dados na tabela atual. Você pode selecionar o número de linhas incluído na amostra ou especificar a porcentagem de observações incluída na amostra. Você também pode especificar uma semente para geração.

Posição dos dados

O campo **Row** mostra o número de linhas no início do conjunto de dados.

Amostra Aleatória

Utilize as setas para cima e para baixo para especificar o número de linhas ou a relação ou proporção a ser incluída na amostra aleatória. Quando você muda um destes itens, os outros se ajustam conseqüentemente. Você pode especificar a semente para usar na aleatorização. Para posições que são menos que ou igual a zero, o tempo do dia é utilizado para inicializar a semente.

Criando Relatórios

Você pode criar um relatório detalhado que lista parte de seus dados, ou você pode criar um relatório em tabelas que resume seus dados.

Para criar um relatório detalhado, selecione **List Data** do menu **Reports**. Para criar uma tabela resumida, selecione **Tables** do menu **Reports**, e escolha um dos cinco estilos de tabela.

5. Transformando Dados

5.1 Diálogo Compute

Selecione **Transform** do menu **Data**, e então selecione **Compute** para mostrar o diálogo Compute. Você pode utilizar o diálogo Compute para especificar uma equação, fórmula ou expressão para criar uma nova coluna na sua tabela de dados. Escreva a expressão na caixa de expressão, ou utilize uma combinação pela tipificação e seleção de variáveis, funções, e operadores. Uma coluna numérica é criada por definição.

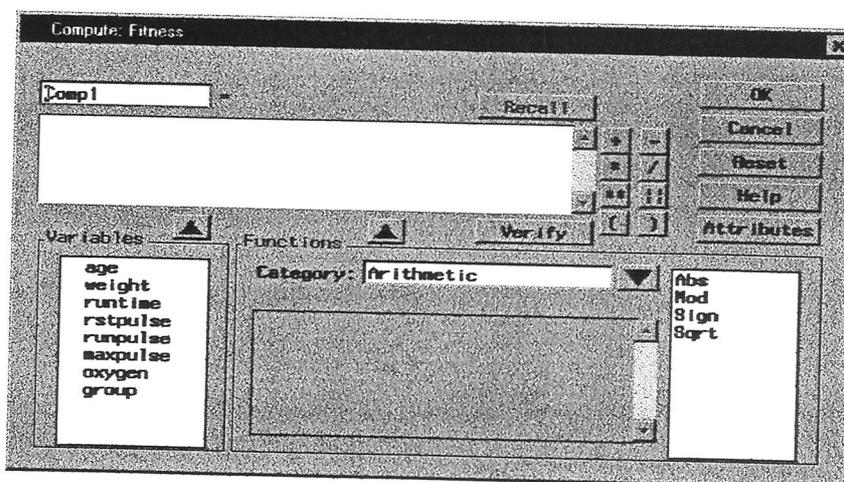


Figura 5.1 Diálogo Compute

Coluna Nome

Por definição, a coluna nome é CompN, onde N é o número mais baixo que produz um único nome. Substitua a coluna nome definida escrevendo um nome de sua escolha.

Botão Recall

Se você já utilizou o diálogo Compute para adicionar uma coluna para a tabela de dados atual, clique no botão **Recall** para preencher a Caixa de Expressão e o diálogo Características das Colunas com o ambiente que define a coluna. O botão **Recall** cria a expressão mais recente.

Caixa de Expressão

A Caixa de Expressão mostra as variáveis, funções, e operadores que compõe a expressão para a coluna computada. Clique em um operador à direita da caixa de expressão para adicionar à expressão, ou escreva o operador na caixa de expressão.

Botão Atributos

O botão Atributos mostra o Column Attributes Dialog, no qual você pode especificar o nome, rótulo e outros atributos para a coluna computada. Se você quer criar uma coluna com valores caracteres, utilize este diálogo para colocar o tipo de variável para o caractere. Numérico é o tipo de variável definida.

Botão Verify

Clique no botão **Verify** para ter certeza que sua expressão é válida. Parâmetros de funções não são verificados e o tipo de variável não é levado em conta.

Variáveis

Clique duplo em uma variável na lista de variáveis ou selecione uma variável e clique na seta para cima na lista para adicionar à sua expressão. Você também pode escrever em um nome de variável.

Funções

Funções são organizadas em categorias. Selecione uma categoria clicando a seta próximo ao campo **Category**. A informação pesquisada sobre uma função selecionando isto. Adicione uma função para a expressão com um clique duplo ou selecionando a função e clicando com a seta. Você também pode escrever em uma função. As funções mostradas são um subcomponentes de todas as funções do SAS.

5.2 Diálogo Rank

Selecione **Transform** do menu **Data**, e selecione **Rank** para mostrar o diálogo Rank.

Lista de Candidatos

Para ordenar variáveis, selecione uma variável da lista de candidatos e clique no botão **Rank**.

Botão BY Group

Selecione uma variável da lista e clique o botão **BY Group** para designá-la como um agrupamento de variáveis. Se você especifica o grupo de variáveis, os postos são calculados separadamente dentro dos níveis definidos pelo grupo de variáveis.

Direção de Ordenação

Selecione **Smallest to largest** ou **Largest to smallest** como sua direção de ordenação.

Botão Options

Para controlar como as variáveis são reordenadas, clique no botão Options.

5.3 Diálogo Standardize

Selecione **Transform** do menu **Data**, e então selecione **Standardize** para mostrar o diálogo Standardize. Utilize este diálogo para padronizar qualquer variável quantitativa.

Lista de Candidatos

Selecione uma variável da lista de candidatas para adicionar às listas Standardize e BY Group.

Botão Standardize

Para padronizar uma variável, selecione uma variável e clique no botão **Standardize**.

Botão BY Group

Para especificar um grupo de variável, selecione a variável e clique o botão **BY Group**. Se você especificar o grupo de variáveis, a padronização é feita separadamente dentro dos níveis definidos pelos grupos de variáveis.

Standardize Para

Por definição, variáveis são padronizadas para ter média zero e desvio padrão um. Você pode escolher outros valores.

Diálogo Recodificação de Valores da Informação

Freqüentemente na execução de uma análise, você pode encontrar que uma certa variável não é exatamente o que você precisa. Por exemplo, você poderia querer ocultar dois ou mais níveis de uma classe de variável no nível um. Valores recodificados permitem que você faça isto.

Selecione **Transform** do menu **Data**, e selecione **Values Recode** para designar as colunas cujos valores você quer substituir.

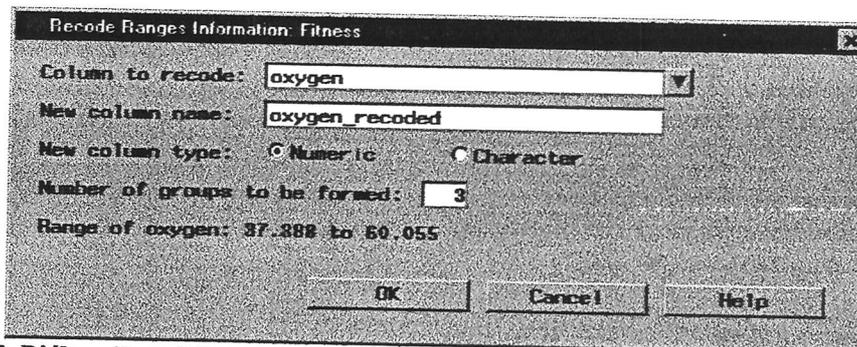


Figura 5.2 Diálogo Recodificação do Valor da Informação

Após você selecionar uma coluna para recodificar, clique **OK** para mostrar um diálogo onde você especifica a recodificação a ser executado.

Coluna para Recodificação

Clique a seta próximo a **Column to recode** para selecionar uma coluna numérica ou de caracter da tabela de dados atual.

Novo Nome da Coluna

O tipo de nome da nova coluna que conterá os novos valores dos dados. A nova coluna tem um nome definido, o qual você pode escrever novamente com o nome de sua escolha.

Novo Tipo de Coluna

O novo tipo de coluna pode ser caracter ou numérico.

Diálogo de Recodificação da Amplitude da Informação

Executando uma análise, você pode descobrir que você gostaria de trabalhar com um fator particular como uma variável de classificação preferentemente do que como uma variável contínua. Recodificação da Amplitude permitem que você crie uma nova variável com níveis distintos baseada nas amplitudes dos valores de uma variável existente.

Selecione **Transform** do menu **Data**, e selecione **Recode Ranges** para designar a coluna cujas amplitudes você quer utilizar.

Após ter selecionado uma coluna para recodificar e o número de grupos que você deseja que a nova variável tenha, clique **OK** para mostrar um diálogo onde você pode especificar a recodificação a ser executada.

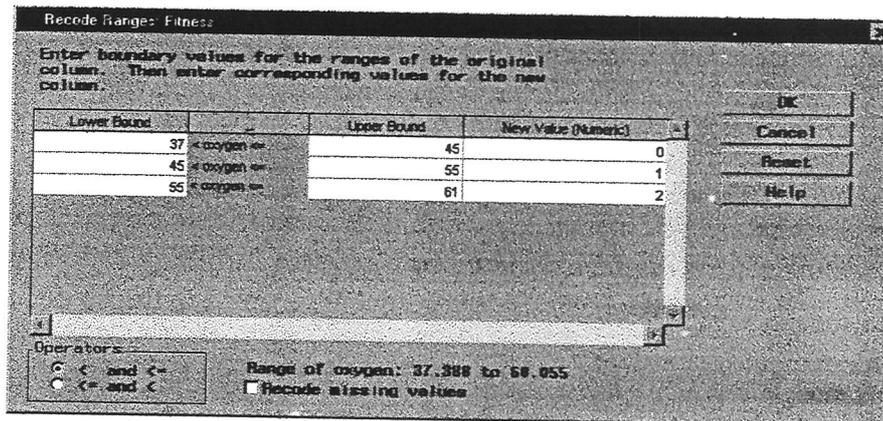


Figura 5.3 Diálogo Recodificação da Amplitude da Informação

Coluna para Recode

Clique a seta próximo a **Column** para selecionar uma coluna numérica da tabela de dados atual.

Novo Nome da Coluna

O tipo de nome da nova coluna que conterá os novos valores dos dados. A nova coluna tem um nome definido, o qual você pode reescrever de acordo com sua escolha.

Novo Tipo de Coluna

O novo tipo de coluna pode ser caracter ou numérico. Se você seleciona **Caracter**, você pode utilizar uma série de letras e/ou sinais para corresponder a cada amplitude.

Numero de Grupos a serem Formados

Você deve especificar o número de grupos em que a amplitude atual será repartida.

Amplitude da coluna nome

Para ajudar você a decidir quantos grupos formar, a amplitude da coluna existente é mostrada no botão deste diálogo.

5.5 Diálogo Tipo Conversão

Selecione **Transform** do menu **Data**, então selecione **Convert Type** para mostrar o diálogo **Convert Type**.

O diálogo **Convert Type** cria novas colunas que contém conversões dos tipos de colunas selecionados para numérico ou caracter.

Lista de Candidatos

A lista de candidatos contém todas as colunas da tabela de dados selecionada. Selecione uma coluna para converter.

Botão To Numeric

Clique o botão **To Numeric** para converter o tipo de coluna selecionada para Numérico. Se o tipo da coluna selecionada já é numérico, este botão estará bloqueado (cinza).

Botão To Character

Clique o botão **To Character** para converter o tipo de coluna selecionada para Character. Se o tipo da coluna selecionada já é character, este botão estará bloqueado (cinza).

Botão Remove

Selecione uma coluna na lista **To Numeric** ou **To Character** e clique o botão **Remove** para remover a coluna da lista de colunas a ser convertida.

5.6 Calculando Logaritmos

Selecione **Transform** do menu **Data**, e selecione **Log(Y)** para calcular o logaritmo natural dos valores na coluna selecionada. Uma nova coluna contendo o logaritmo de cada valor é criada.

5.7 Calculando Raízes Quadradas

Selecione **Transform** do menu **Data**, e selecione **Sqrt(Y)** para calcular a raiz quadrada dos valores na coluna selecionada. Uma nova coluna contendo a raiz quadrada de cada valor é criada.

5.8 Calculando Recíprocas

Selecione **Transform** do menu **Data**, e selecione **1/Y** para calcular a recíproca dos valores na coluna selecionada. Uma nova coluna contendo a recíproca de cada valor é criada.

5.9 Calculando Quadrados

Selecione **Transform** do menu **Data**, e selecione **Y*Y** para calcular o quadrado dos valores na coluna selecionada. A nova coluna contendo o quadrado de cada valor é criada.

5.10 Calculando Expoentes – Construindo Exponenciais

Selecione **Transform** do menu **Data**, e selecione **Exp(Y)** para construir a constante de e para os valores na coluna selecionada. Uma nova coluna é criada que contém os valores resultantes.

Exercício 6:

Neste exercício você deve efetuar algumas transformações no conjunto de dados que contém informações sobre o peso, altura e comprimento e largura de caixas.

1. Abra o arquivo **medidas.sas**;
2. Certifique-se que o seu banco de dados esteja no modo **Edit**.
3. Clique sobre o nome da variável **Peso**.
4. Selecione a opção **Transform** do menu **Data** e depois clique sobre **Log (Y)** para calcular o logaritmo do peso.

Agora, você deve calcular o volume da caixa.

1. Selecione a opção **Transform** do menu **Data** e depois clique em **Compute**.

2. No primeiro campo, escreva **Volume**.
3. Clique, agora, na variável **Altura**, localizada na caixa **Variables** e em seguida no asterisco que está à direita.
4. Repita este procedimento para a variável **Comp** e por último insira a variável **Largura**.
5. Clique em **OK** para executar o cálculo.

6. Combinando Tabelas

6.1 Diálogo Unindo Tabelas por Colunas

Use esta tarefa para unir tabelas horizontalmente (lado-a lado). Selecione **Combine Tables** do menu **Data** e então selecione **Merge By Columns**. No diálogo Unindo Tabelas por Colunas, você pode selecionar tabelas de dados para unir variáveis que você manterá na tabela consolidada.

Tabelas

Você pode unir até seis tabelas. Escreva o nome da tabela no campo **Table name**, clique na seta para selecionar o nome da tabela no SAS, ou clique no botão **Browse** para selecionar o arquivo a partir do seu diretório.

Botão More

Clique no botão **More** para unir mais que duas tabelas.

Tabelas combinadas serão guardadas

Você pode escolher se as novas combinações serão mostradas somente pelas linhas combinadas, linhas que se adaptam àquela na tabela1 ou todas as linhas. Para estas três opções **Merge By Variables** pode ser usado.

Unindo Variáveis

Variáveis comuns entre as tabelas que foram escolhidas para serem combinadas são listadas em **Common Variables list**.

Botão Merge By

Selecione uma variável comum e clique no botão **Merge By** para adicioná-la a lista de **Merge By Variables**.

Selecione uma variável e clique no botão **Remove** para removê-la da lista **Merge By Variables**.

Botão Variable

Clique no **botão Variable** para selecionar as variáveis que você quer mostrar na sua tabela de variáveis unidas.

6.2 Diálogo Concatenar Tabelas por Linhas

Use esta tarefa para combinar tabelas verticalmente (uma no topo da outra). Selecione **Combine Tables** do menu **Data**, e então selecione **Concatenate by Rows**. No diálogo Concatenar Tabelas por Linhas você pode selecionar as tabelas de dados e anexar as linhas ou intercalar pelas variáveis que você selecionou.

Tabelas para Concatenar

Clique no botão **Open SAS Data** para abrir a tabela de dados SAS. Clique no botão **Browse** para selecionar o arquivo do seu diretório de sistema de operação.

Para mudar a ordem das tabelas que você anexou, selecione a tabela e clique na linha superior ou inferior para mover a tabela um nível acima ou abaixo na lista.

Para remover a tabela da lista, selecione a tabela e clique no botão **Remove**.

Métodos de Concatenação

Selecione **Append** para anexar as tabelas que você tem selecionadas. Se você escolheu anexar tabelas, você pode mudar a ordem das tabelas na lista. O método de anexar empilha o conjunto de dados um no topo do outro.

Selecione **Interleave** para intercalar as tabelas. Este método separa o conjunto de dados combinado e intercala variáveis.

Intercalar Variáveis

Variáveis comuns entre as tabelas escolhidas para concatenar são listadas na lista de Variáveis Comuns. Selecione a variável comum e clique no botão **Interleave By** para adicioná-la a lista das variáveis intercaláveis. Selecione a variável e clique no botão **Remove** para removê-la da lista de Variáveis intercaláveis.

Botão Variáveis

Clique no botão **Variables** para selecionar as variáveis que você quer que sejam mostradas no seu conjunto de dados.

6.3 Exemplo

Neste exemplo, você combina colunas selecionadas de dois conjunto de dados e as edita em uma nova tabela.

Cada conjunto de dados contém os resultados de testes de sabores de cereais matinais. Cada cereal é avaliado por vários juízes, na escala de 1 a 5. Depois que você concatenar os dois conjunto de dados, você vai dividi-lo em colunas de avaliação e por número da amostra.

Abrindo o conjunto de dados para Edição

Para selecionar o conjunto de dados e trazê-los para a nova tabela de dados do Analyst, siga as seguintes etapas:

1. Selecione **Tools Sample Data...**
2. Selecione *JRating1* e *JRating*.
3. Clique em **OK** para criar uma amostra do conjunto de dados em seu diretório Sasuser.
4. Selecione **Data Combine Tables Concatenate By Rows...**
5. Clique no botão **Open SAS Data**. Selecione Sasuser da lista de biblioteca. Selecione *Jrating1* da lista de membros e clique em **OK**.
6. No diálogo concatenar tabelas por linhas, clique novamente no botão **Open SAS Data**. Selecione Sasuser da lista de bibliotecas. Selecione *Jrating2* da lista de membros e clique em **OK**.

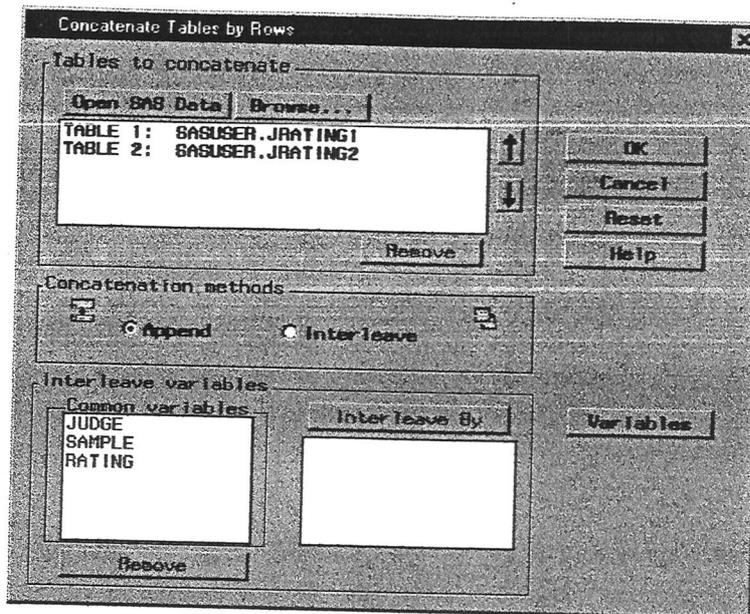


Figura 6.1 Diálogo Concatenar Tabelas por Linhas

7. Selecione **Interleave**.
8. Selecione **JUDGE** e **SAMPLE** da lista de **Common variables** e clique no botão **Interleave By** para usar **JUDGE** e **SAMPLE** como variáveis pelas quais as linhas dos dados serão combinadas

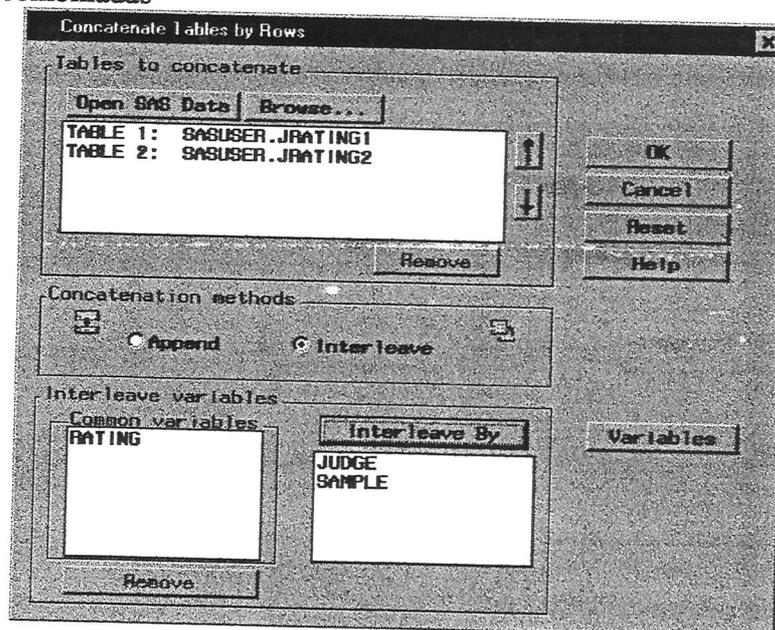


Figura 6.2 Combinar por Variáveis Comuns

9. Clique no botão **Variables** para selecionar as colunas e incluir em uma nova tabela de dados.

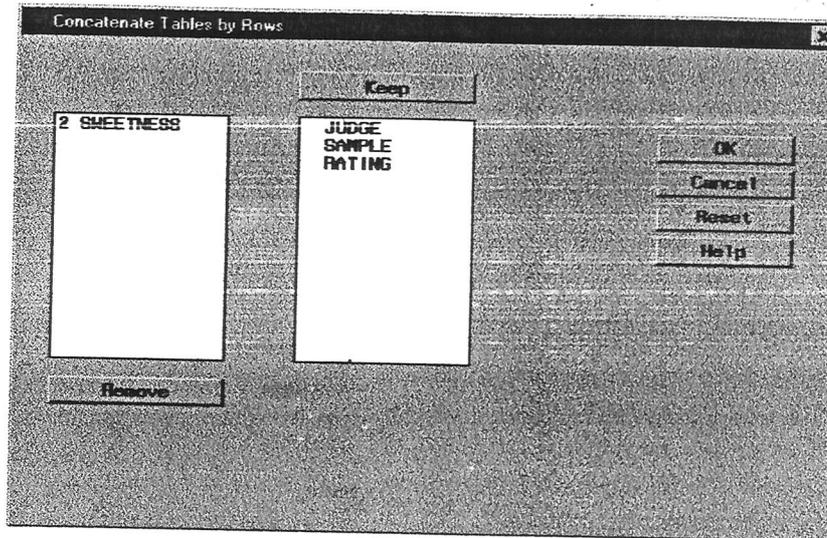


Figura 6.3 Selecionar colunas de uma nova Tabela de Dados.

Somente aquelas colunas comuns para ambas tabelas de dados são mantidas por padrão, como mostrada na lista **Keep**. A coluna SWEETNESS não é mantida como parte da tabela resultante. O número que precede o nome da coluna SWEETNES representa a tabela de dados a que esta variável pertence.

10. Clique em **OK** para retornar para o Diálogo Concatenar Tabelas por Linhas. Clique em **OK** novamente para exibir a nova tabela de dados combinada em uma janela de resultados.

	Judge	Sample	Rating	
1	A23	1	5	
2	A23	2	4	
3	A23	3	4	
4	A23	4	2	
5	B37	1	4	
6	B37	2	3	
7	B37	3	3	
8	B37	4	1	
9	C12	1	5	
10	C12	2	4	
11	C12	3	2	
12	C12	4	3	
13	D77	1	3	
14	D77	2	5	
15	D77	3	4	

Figura 6.4 Tabela Combinada

11. Para modificar uma tabela combinada, você precisa abri-la na tabela de dados do Analyst. Feche a janela de resultados. Selecione o nó **Combine Table** na árvore do projeto e clique com o botão direito do mouse para exibir o menu desdobrável.. Selecione **Open**.

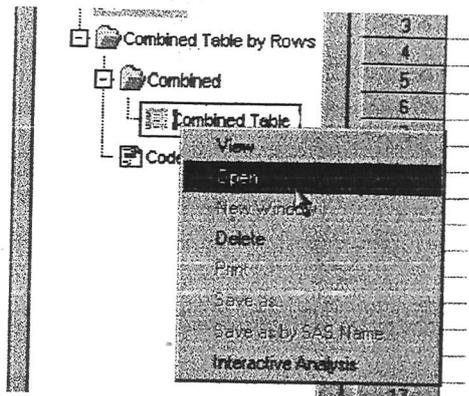


Figura 6.5 Abrir a Tabela Combinada.

12. Por padrão, as tabelas de dados são abertas no modo Browse. Selecione **Edit Mode Edit** para mudar o modo de Browse para Edit.

Modificando os Dados

Em uma tabela de dados você pode modificar os dados dividindo colunas dessa forma uma nova coluna é gerada quando os valores das variáveis mudam. Você pode também subdividir dentro das amplitudes.

Para subdividir os dados por amplitudes e dividir colunas de acordo com o número da amostra, siga as seguintes etapas:

1. Divida o resultado do teste dos sabores em três categorias: Bom, Mediano e Ruim. Selecione **Data Transform Recode Ranges...**
2. Clique na seta para **Column to recode** e selecione Rating. Escreva **taste_test** no campo **New column name**. Mude **New Column type** para **Character**. Escreva **3** no campo **Number of groups to be formed** para designar três amplitudes para testes de sabores.

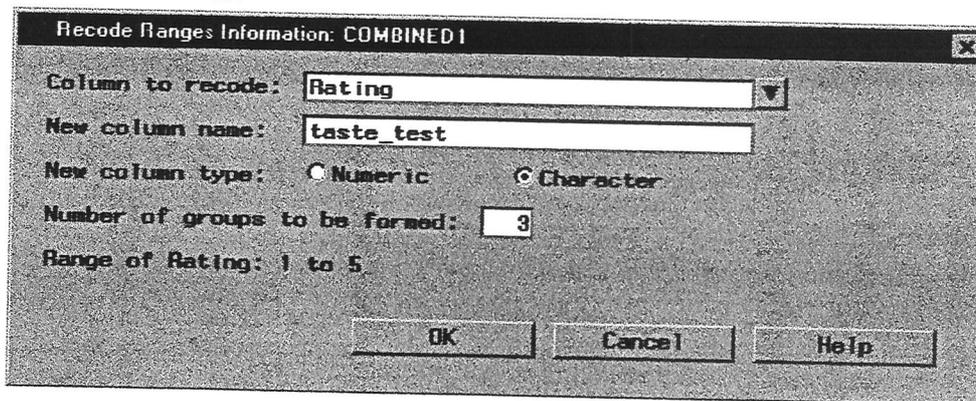


Figura 6.6 Diálogo Recodificação da Amplitude da Informação

Clique em **OK** para especificar as novas amplitudes.

3. Na primeira linha, escreva **0** na coluna **Lower Bound** e **2** na coluna **Upper Bound**. Escreva **Bad** na coluna **New Value**.
4. Quando você pressionar a tecla Enter, o valor do limite superior da linha anterior é automaticamente preenchido com o limite inferior da linha corrente. Escreva **3** na coluna **Upper Bound** e **Mediocre** na coluna **New Value**.
5. Mova seu cursor para a terceira linha. Escreva **5** na coluna **Upper Bound** e **Good** na janela **New Value**.

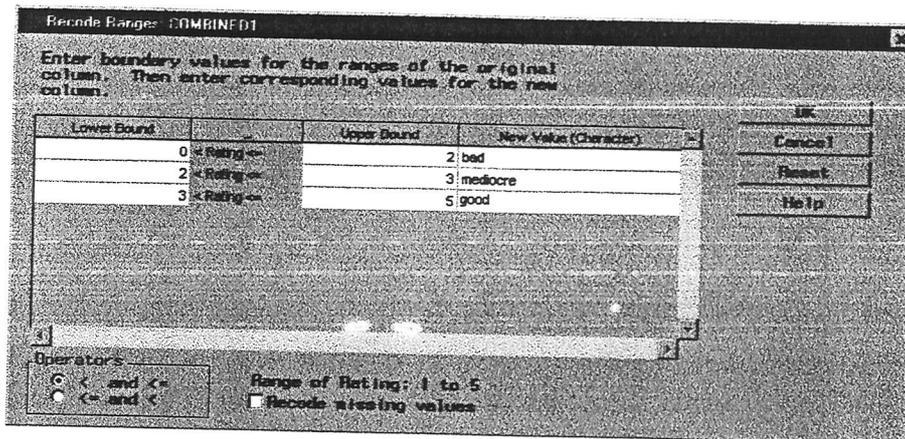


Figura 6.7 Valores Limites

6. Clique em **OK** para salvar seus novos valores limites. Na nova tabela, suas amplitudes são exibidas na coluna `taste_test`.

	Sample	Rating	taste_test
1	1	5	good
2	2	4	good
3	3	4	good
4	4	2	bad
5	1	4	good
6	2	3	mediocre
7	3	3	mediocre
8	4	1	bad
9	1	5	good
10	2	4	good
11	3	2	bad
12	4	3	mediocre
13	1	3	mediocre
14	2	5	good

Figura 6.8 Tabela com a coluna `taste_test`

7. Remova a coluna **Rating** das colunas selecionadas e selecione **Delete** do menu desdobrável. Clique em **OK** no diálogo Deletar Itens.

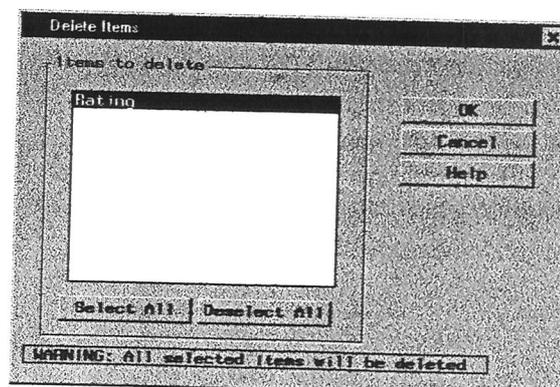


Figura 6.9 Deletando a Coluna Rating

8. Você está dividindo a coluna `taste_test` pela coluna `Sample` assim, o teste de sabor para cada amostra é exibido por juiz. Selecione **Data Spit Columns**.

9. No diálogo Dividir Colunas, selecione taste_test da lista e clique no botão **Split Column**. Selecione Sample da lista e clique no botão **Split By**.
10. Selecione **User-defined names** para a coluna de nomes. Escreva **Sample_** no campo **Column name prefix**.

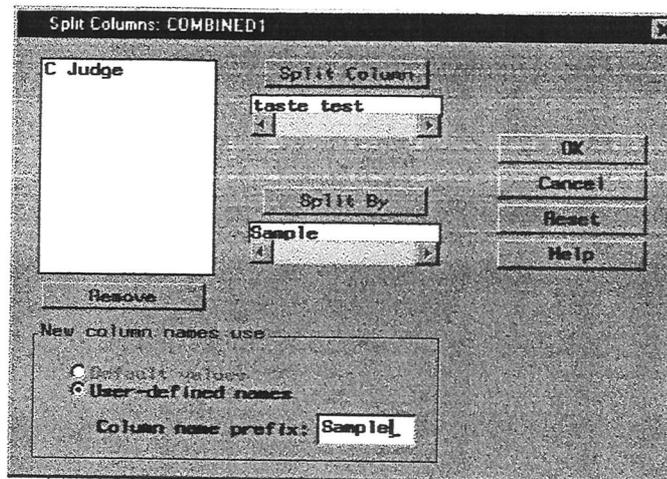


Figura 6.10 Dividir a Coluna Taste_test por Amostra

11. Clique em **OK**. A tabela resultante exibe os resultados do teste de sabor para cada participante do júri.

	Judge	Sample_1	Sample_2	Sample_3	Sample_4
1	A22	good	good	good	bad
2	B37	good	med iocre	med iocre	bad
3	C12	good	good	bad	med iocre
4	D77	med iocre	good	good	bad
5	E48	good	good	med iocre	med iocre
6	R22	good	good	med iocre	bad
7	S69	good	med iocre	bad	bad
8	T15	med iocre	good	med iocre	med iocre
9	U86	good	med iocre	med iocre	bad
10	V83	good	med iocre	good	med iocre
11	W91	med iocre	bad	med iocre	bad
12	X08	good	good	good	bad

Figura 6.11 Tabela com Colunas Divididas

Exercício 7:

Você precisa concatenar dois arquivos com informação da quantidade de chuva em algumas cidades. Para prosseguir com as análises você precisa concatená-los:

- 1 Abra o arquivo **idades1.xls** existente no seu diretório.
- 2 Clique em **Combine Tables** do menu **Data** e selecione a opção **Concatenate By Rows**.
- 3 Clique no botão **Browse** e procure o arquivo **idades2.xls** contido no seu diretório.
- 4 Clique em **OK** para importar os dados para o SAS e em seguida clique em **OK** novamente para executar a tarefa.
- 5 Salve a nova tabela de dados no seu diretório.

7. Trabalhando com colunas

7.1 Diálogo Mover Colunas

Você pode mover colunas selecionando uma ou mais colunas e selecionando **Move** do menu desdobrável, ou selecionando **Columns** do menu **View**, e então selecionando **Move**. Para mover uma coluna destaca-se na lista **Column Descending** sobre a condução **Alphabetical**.

Selecione **Save order with data** para salvar esta ordem com os dados na tabela de dados.

Clique no botão **OK** quando as colunas estão na ordem desejada.

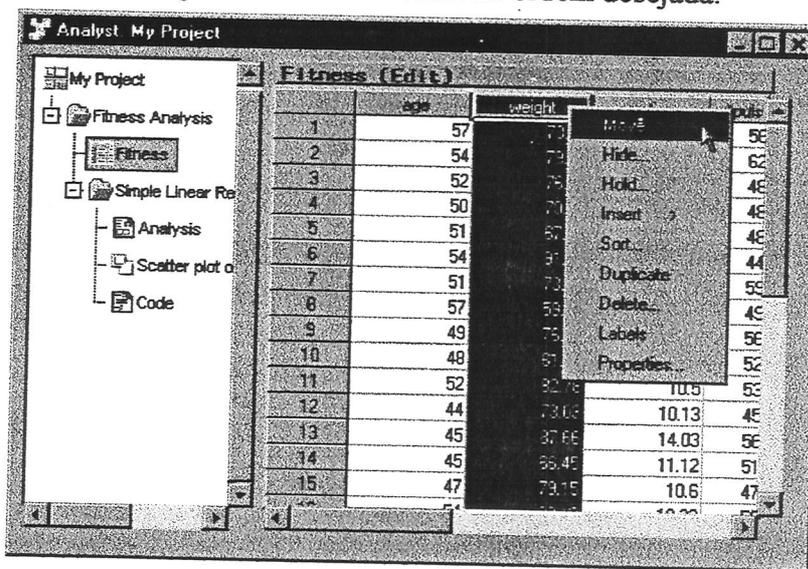


Figura 7.1 Menu Desdobrável da Coluna

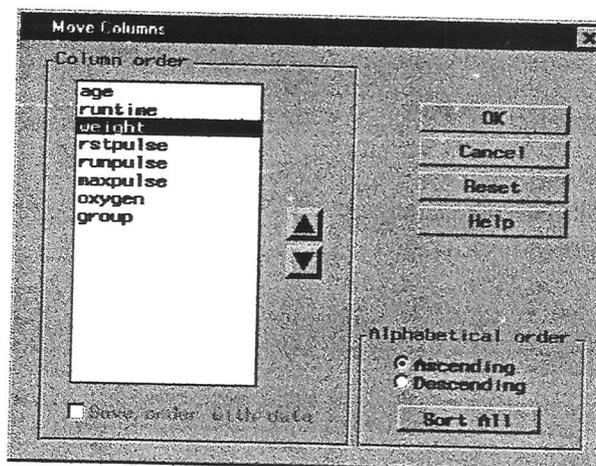


Figura 7.2 Diálogo Mover Colunas

7.2 Diálogo Escondendo Colunas

Para esconder uma coluna ou colunas da exibição, selecione as colunas e selecione **Hide** do menu desdobrável, ou selecione **Columns** do menu **View**, e então selecione **Hide**. Colunas escondidas ainda aparecem na lista de variáveis a menos que você especifique-as para ser excluídas.

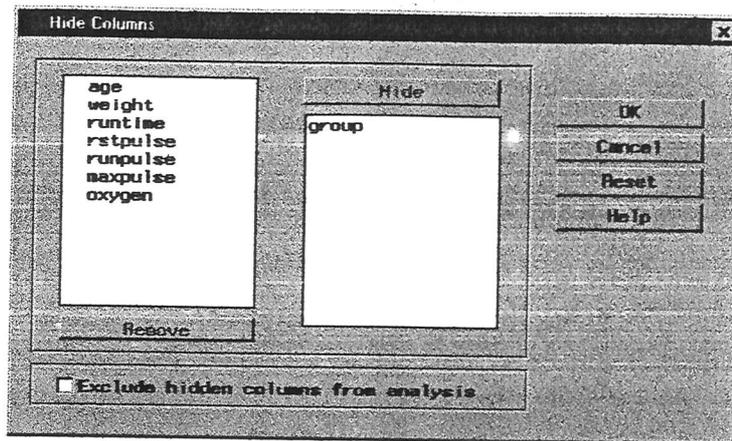


Figura 7.3 Diálogo Esconder Colunas

Lista de Candidatos

A lista de candidatos contém as colunas na tabela. Para esconder as colunas, selecione a coluna desejada e clique no botão **Hide**.

Botão Remove

Para reexibir colunas, selecione a coluna desejada e clique no botão **Remove**.

Botão Hide

Clique no botão **Hide** para adicionar uma coluna selecionada na lista de colunas escondidas.

Excluir Colunas Escondidas da Análise

Selecione esta caixa para especificar que colunas escondidas não são disponíveis nesta análise.

7.3 Diálogo Organizar Colunas

O diálogo Organizar Colunas organiza a coluna selecionada e todas as colunas para a esquerda no lugar enquanto você lista através das colunas para a direita da coluna organizada. Selecione **Columns** do menu **View**, e então selecione **Hold** para exibir o Diálogo Hold ou selecione a coluna e selecione **Hold** do menu desdobrável.

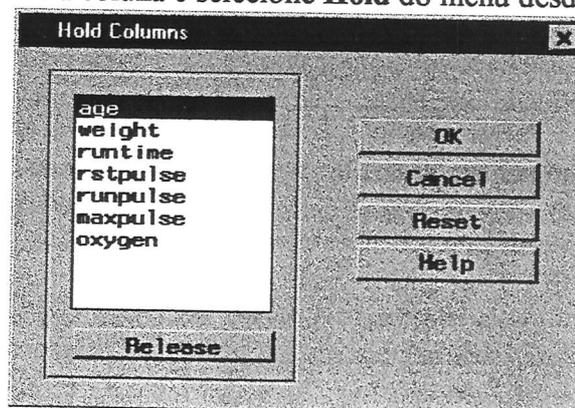


Figura 7.4 Diálogo Organizar Colunas

Lista de Colunas

Selecione uma coluna da lista de colunas e clique em **OK** para organizá-la.

Botão Realease

Selecione uma coluna preparada da lista de colunas e clique no botão **Realease** para resolvê-la.

7.4 Inserindo Colunas

Para inserir uma ou mais colunas, selecione uma coluna e selecione **Insert** do menu, ou selecione **Insert Columns** do menu **Edit**. Então selecione o texto da coluna **Character** ou **Numeric**. A nova coluna é inserida na esquerda da coluna selecionada. Se você selecionar mais que uma coluna, colunas iguais ao número que você selecionou são inseridas na esquerda desta primeira coluna.

7.5 Duplicando Colunas

Para duplicar uma ou mais colunas, selecione uma coluna e selecione **Duplicate** no menu. A coluna duplicada é inserida a esquerda da coluna selecionada. Se você selecionar mais que uma coluna, um número igual de colunas que você selecionou são duplicadas a esquerda da primeira selecionada.

7.6 Diálogo Deletar Itens

Selecione uma ou mais colunas e selecione **Delete** do menu para exibir o diálogo Deletar Itens. Selecione a coluna que você quer deletar e clique em **OK**. Clique no botão **Select All** para deletar todas as colunas na lista. Clique no botão **Deselect All** para impedir que as sejam deletadas.

7.7 Diálogo Empilhar Colunas

Selecione **Stack Columns** do menu **Data** para exibir o diálogo Stack Columns. Este diálogo cria um novo conjunto de dados empilhando as colunas especificadas dentro de uma coluna comum. Os valores nas outras colunas são preservados em um novo conjunto de dados. A coluna de origem no novo conjunto de dados contém o nome da coluna o conjunto de dados original que contém os valores empilhados. Empilhar colunas pode ser particularmente usado quando uma medida é repetida muitas vezes dentro de cada observação, por exemplo, medida de pressão sanguínea tomada em várias visitas a uma clínica. A saída do conjunto de dados para cada medida de pressão sanguínea apareceria em observações separadas identificadas por uma variável de origem que indica a visita.

Lista de Candidatos

A lista de candidatos exibe todas as colunas na tabela de dados. Você pode selecionar as colunas desta lista para empilhar. Você não pode unir tipos de colunas quando empilhar, você pode empilhar colunas de caracteres com caracteres e numéricas com numéricas. Note que selecionando variáveis não contíguas da lista de candidatos, somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muito locais você pode então pressionar **Ctrl** enquanto clica nas variáveis.

Botão Remove

O Botão **Remove** Habilita você a remover colunas de qualquer lista exceto da lista de candidatos, que lista todas as colunas da tabela de dados. Selecione as colunas, então clique

no botão **Remove**, e as colunas serão removidas da lista. Você pode também remover as colunas da lista dando um duplo clique sobre elas.

Botão Stack

Selecione as colunas da lista de candidatos e clique no botão **Stack** para selecionar as colunas para empilhá-las.

Nomes das Novas Colunas

Você pode usar nomes padrões ou símbolos nos novos nomes para as colunas empilhadas e de origem. A coluna de origem identifica a coluna original de cada valor empilhado.

7.8 Diálogo Dividir Colunas

Você pode dividir uma coluna para produzir novas colunas para qualquer valor das variáveis modificadas. Selecione **Split Columns** do menu **Data** para exibir o diálogo dividir Colunas.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis da tabela de dados. Você pode selecionar a partir desta lista para exibir como uma coluna dividida e a variável pela qual será dividida.

Botão Remove

O botão **Remove** habilita você a remover colunas de qualquer lista, exceto da lista de candidatos, que lista todas as colunas da tabela de dados. Selecione as colunas, então clique no botão **Remove**, e as colunas serão removidas da lista. Você também pode remover colunas da lista dando um duplo clique sobre elas.

Botão Dividir Coluna

Selecione uma coluna da lista de candidatos e clique no botão **Split Column** para designar que coluna dividir.

Botão Split By

Selecione uma variável da lista de candidatos e clique no botão **Split By** para designar uma variável para dividir a primeira coluna.

Uso do Nome da Nova Coluna

Você pode usar nomes padrões ou símbolos nos novos nomes para colunas divididas ou de origem, se o tipo da coluna divididas for caracter. Colunas numéricas não tem nomes padrões.

7.9 Diálogo Mudar o Lugar (Transpose)

Selecione **Transpose** do menu **Data** para mudar o lugar ou posição dos dados na tabela.

Lista de Candidatos

A lista de candidatos exibe variáveis do conjunto de dados.

Botão Transpose

Selecione um nome da variável e clique no botão **Transpose** para adicioná-la a lista de variáveis a serem mudadas de ordem ou posição.

Botão Group By

Selecione um nome da variável e clique em **Group By** para adicioná-la a lista de agrupá-lá Por. Esta lista contém as variáveis a serem usadas para agrupar as variáveis na transposição do conjunto de dados. Conjuntos separados de observações são criados para cada nível ou combinação de níveis de variáveis agrupadas.

Nomes das Novas Colunas

Você pode usar valores das variáveis nos dados originais nos conjunto de dados transposto.

Especifique a coluna original e o nome da coluna prefixada a ser usado para os nomes das variáveis dos dados transformados.

8. Construindo Gráficos

8.1 Diálogo Gráfico de Barras

Gráfico de barras Horizontal e vertical são exibidos em seu conjunto de dados na forma de gráfico de barras bidimensionais ou tridimensionais. Selecione **Bar Chart** do menu **Graphs**, e então selecione **Horizontal** ou **Vertical** para exibir seus dados em um gráfico de barras.

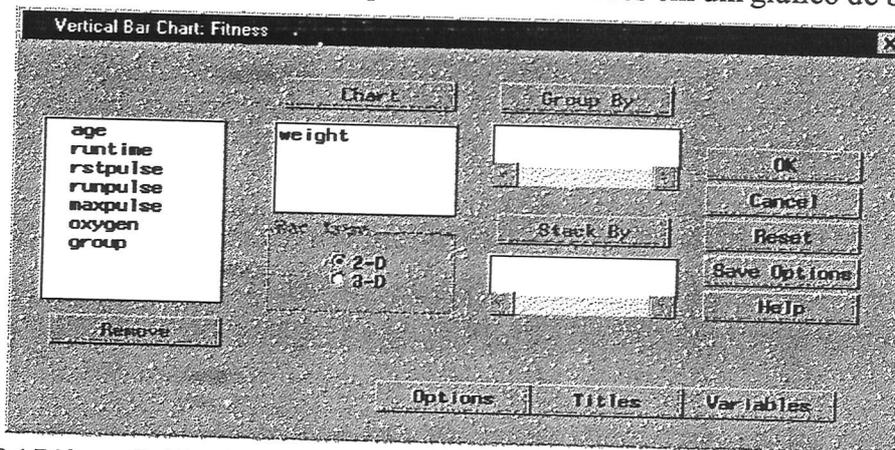


Figura 8.1 Diálogo Gráfico de Barras Verticais

Lista de Candidatos

A lista de candidatos mostra todas as variáveis do conjunto de dados. Você pode selecionar variáveis da lista para exibi-las em um gráfico de barras. Note que selecionando variáveis não contíguas da lista de candidatos, somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais, você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão Chart

Selecione as variáveis da lista de candidatos e clique no botão **Chart** para incluir a variável no gráfico de barra.

Tipo de Barra

Selecione 2-D ou 3-D sobre o Tipo de Barra para especificar se você quer exibir gráficos bidimensionais ou tridimensionais.

Botão Group By

Selecione uma variável da lista de candidatos e clique no botão **Group By** para adicioná-la nas variáveis agrupadas no gráfico de barras.

Botão Stack By

Selecione uma variável da lista de candidatos e clique no botão **Stack By** para adicioná-la e usá-la como uma variável empilhada no gráfico de barras.

Opções do Gráfico de Barras

As seguintes opções estão disponíveis para ajustar o seu gráfico de barras.

- **Options:** Controla a aparência do gráfico de barras.

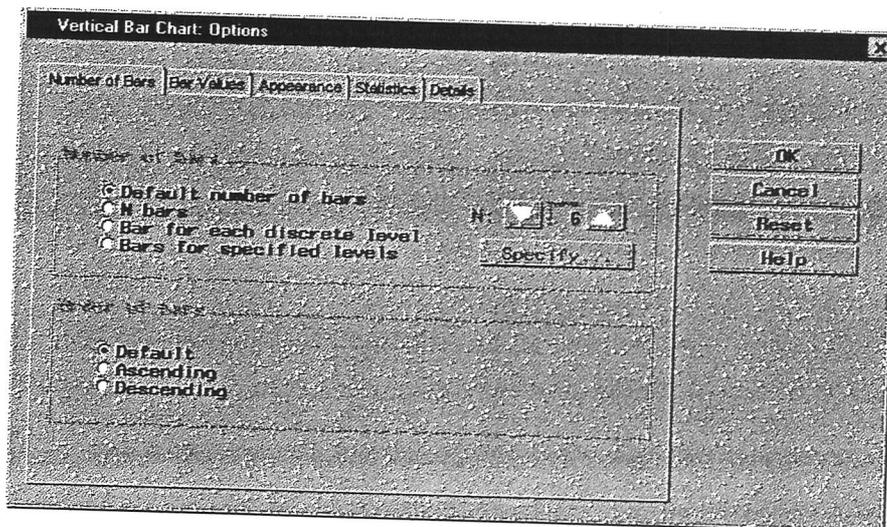


Figura 8.2. Placa Número de Barras

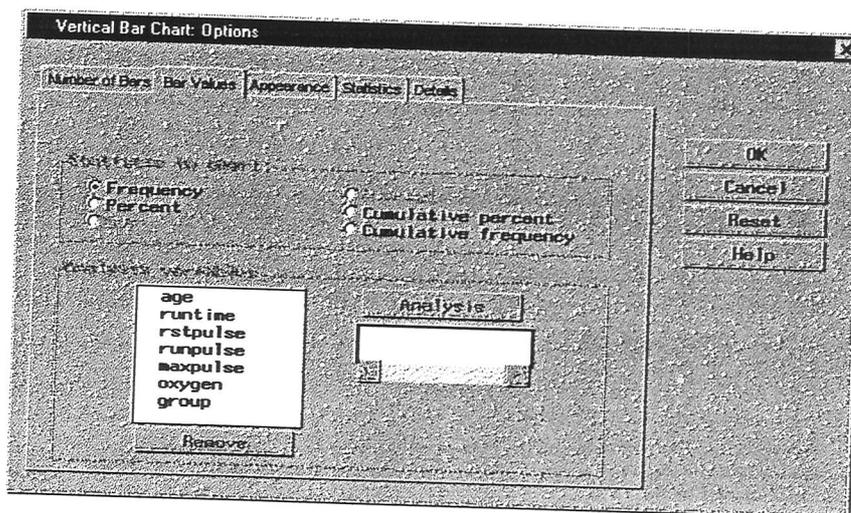


Figura 8.3. Placa Valores da Barra

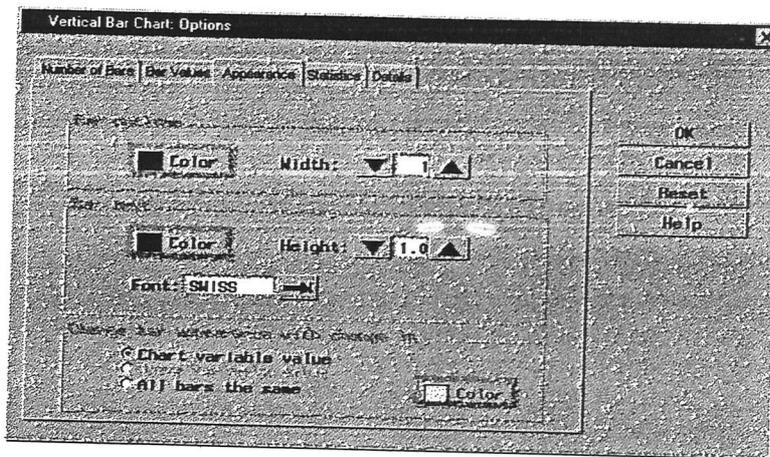


Figura 8.4. Placa Aparência

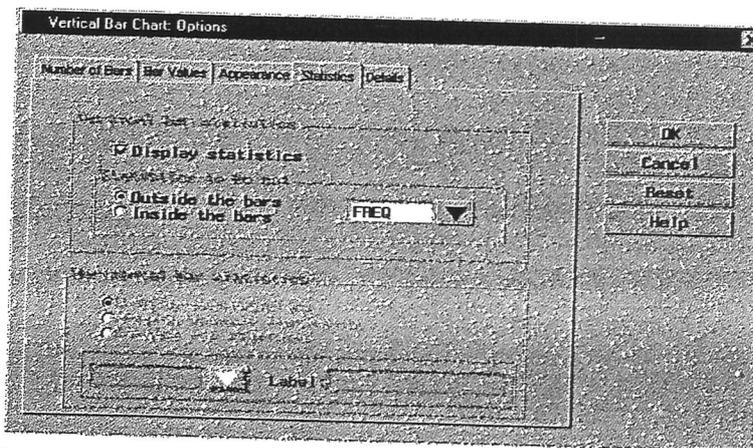


Figura 8.5. Placa Estatísticas

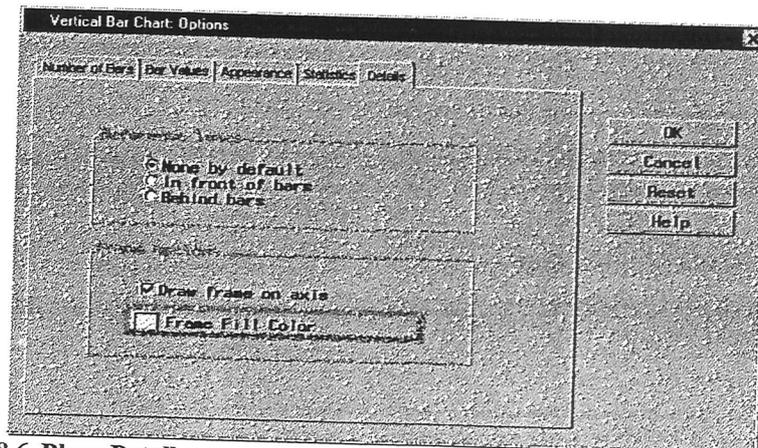


Figura 8.6. Placa Detalhes

- Títulos: Especifica os títulos dos resultados.

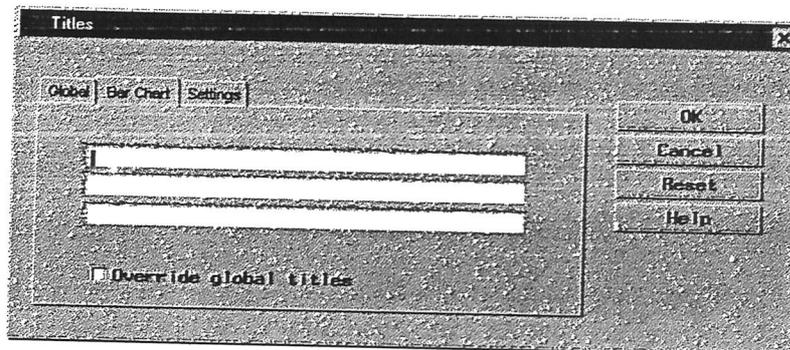


Figura 8.7. Diálogo Títulos, Placa Diagrama de Barras

- Variáveis: Especifica POR grupos de variáveis.

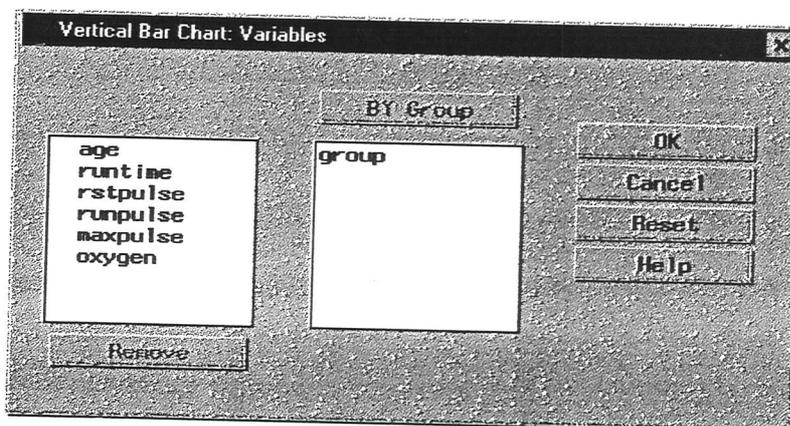


Figura 8.8. Gráfico de Barras Verticais: Diálogo Variáveis

8.1.1 Exemplo:

Neste exemplo, você cria um gráfico de barras usando um conjunto de dados de Aptidão Física.

1. Selecione **Tools Sample Data**
2. Selecione **Fitness**
3. Clique em **OK** para criar uma amostra do conjunto de dados em seu diretório Sasuser.
4. Selecione **File Open By SAS Name**.
5. Selecione Sasuser da lista de **bibliotecas**.
6. Selecione Fitness da lista de membros.
7. Clique em **OK** para trazer conjunto de dados Fitness para dentro da sua tabela de dados.

Especificando o gráfico e agrupando as variáveis

Para criar um gráfico de barras verticais em 3-D que compara grupos experimentais e que calcula a média da quantidade de consumo de oxigênio dado o tempo que leva para correr 1.5 milhas, siga as seguintes etapas:

1. Selecione **Graphs Bar Chart Vertical** para exibir o diálogo do gráfico de Barras Vertical.
2. Selecione *runtime* (tempo de corrida) da lista de candidatos e clique em **Chart** para marcar os minutos para correr 1.5 milhas na variável que será representada no gráfico.
3. Sobre **Bar type**, selecione **3-D** para fazer um gráfico de barras tridimensional.
4. Para comparar entre os grupos experimentais, selecione *group* da lista de candidatos e clique em **Group By**.

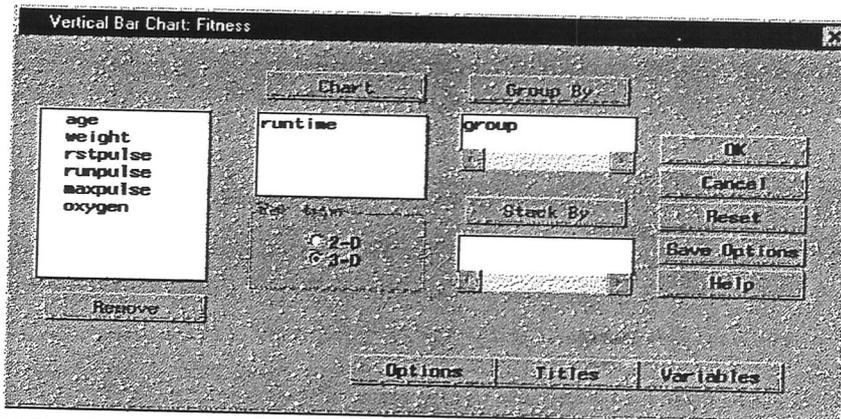


Figura 8.9 Variáveis das Barras e de Agrupamento

Especificando as opções do gráfico de barras

Para especificar seu gráfico de barras, assim como o número e aparência das barras, siga as seguintes etapas:

1. Clique no botão **Options** para exibir o diálogo de Opções do Gráfico de Barras.
2. Sobre **Number of Bars**, selecione **N bars**, e clique na seta para baixo até $N=3$. Pois uma variável de agrupamento foi especificada, barras para os três pontos centrais de *runtime* são exibidas para cada valor do grupo experimental.

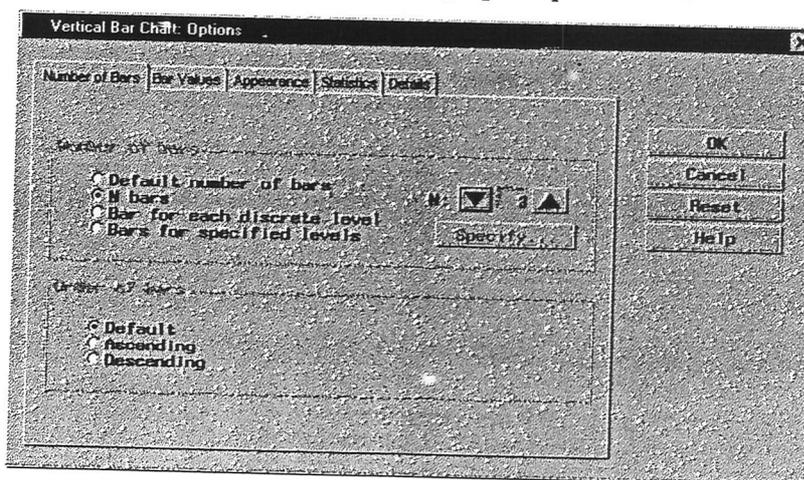


Figura 8.10 Número de Barras

3. Selecione a placa **Bar Values**. Sob **Analysis variables**, selecione *oxygen* da lista de candidatos e clique no botão **Analysis** para fazer o consumo de oxigênio sua variável de análise.
4. Sob **Statistic to chart**, selecione **Average** para exibir o consumo médio de oxigênio por tempo de corrida.

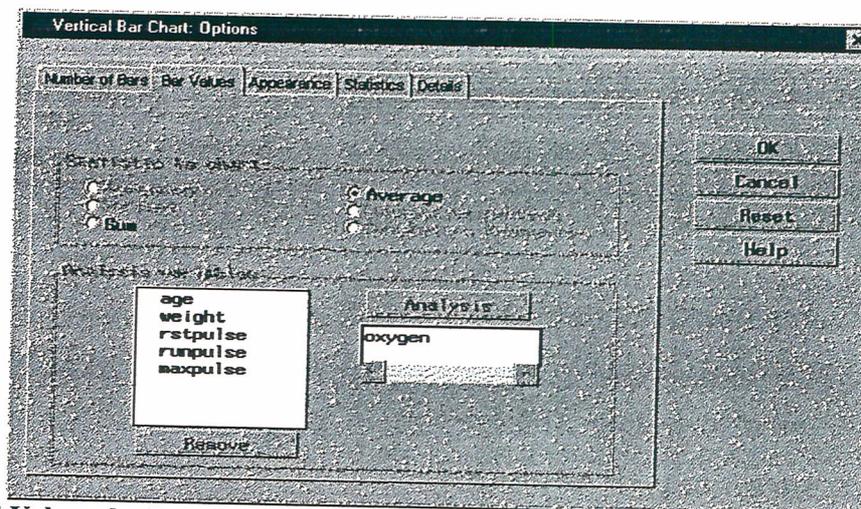


Figura 8.11 Valores das Barras

5. Selecione a placa **Appearance**. Sob **Bar outline**, clique no botão **Color**. Selecione **White** da lista de Color Attributes para deixar o contorno da barra branco.

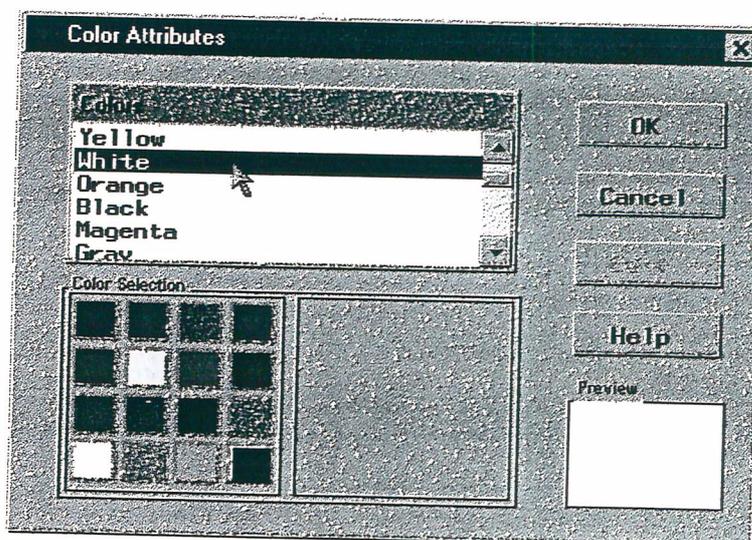


Figura 8.12 Contorno das Barras

Clique em **OK** para fechar a janela de Atributo das Cores e retornar para o diálogo de Opções do Gráfico de Barras.

6. Ainda sobre a placa **Appearance**, selecione **Group variable value** sob **Change bar appearance with change in**.

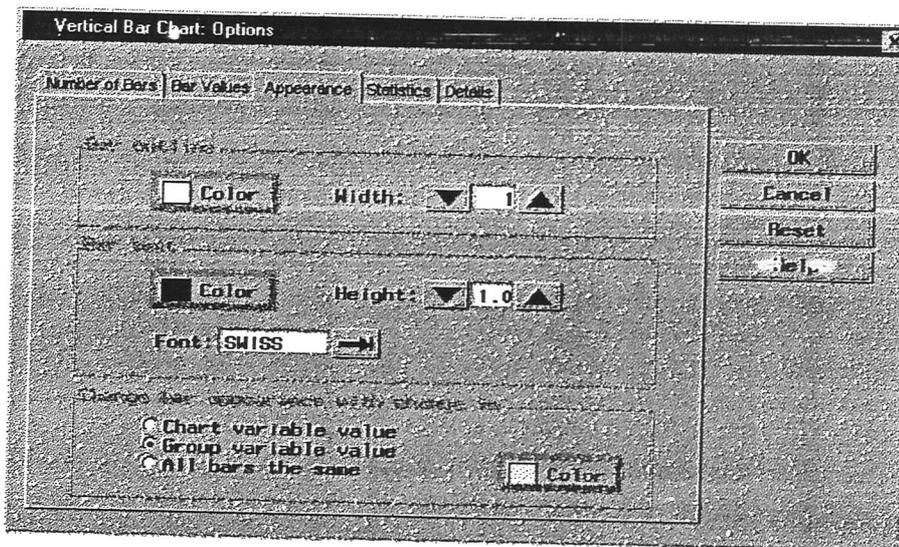


Figura 8.13 Aparência da Barra

7. Clique em **OK** para retornar ao diálogo Gráfico de Barras Verticais.

Especificando Títulos do Gráfico de Barras

Para especificar os títulos para o seu gráfico de barras, siga as seguintes etapas:

1. Clique sobre o botão **Titles** no diálogo de Gráfico de Barras Verticais.
2. Na placa **Bar Chart**, escreva **Runtime and Oxygen Consumed** (Tempo e Oxigênio consumido) no primeiro campo.

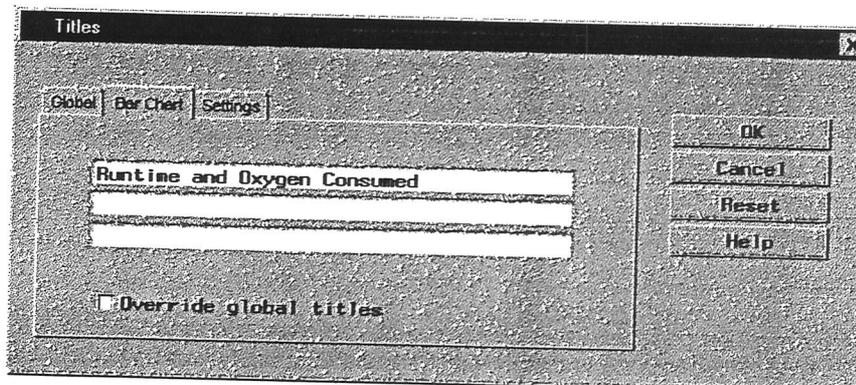


Figura 8.14 Título do Gráfico de Barras.

3. Clique na placa **Global**. Escreva **Fitness Report** (Relatório de Aptidão Física) no primeiro campo. Este título global está salvo entre todas as sessões do Analyst até você mudá-lo.

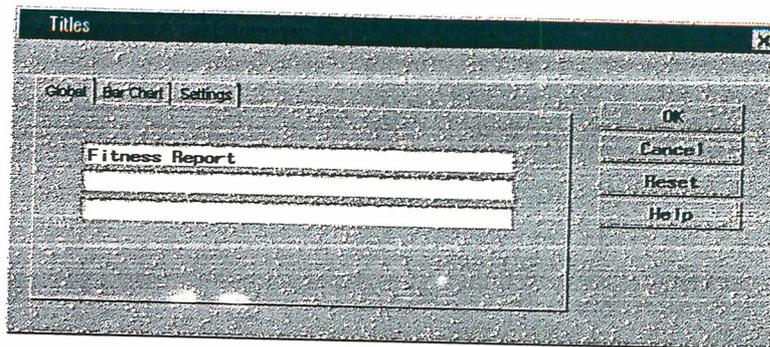


Figura 8.15 Título Global

4. Clique em **OK** para salvar suas mudanças nos títulos.

Gerar um Gráfico de Barras

Para exibir seu gráfico de barras, clique em **OK** no diálogo de Gráfico de Barras Verticais.

Como esperado, grandes quantias de oxigênio são consumidas por corredores mais rápidos. O grupo experimental não se mostra afetado com esta e relacionamento ou quantias médias de consumo de oxigênio. Nenhum membro do grupo experimental 2 esteve entre os corredores mais lentos.

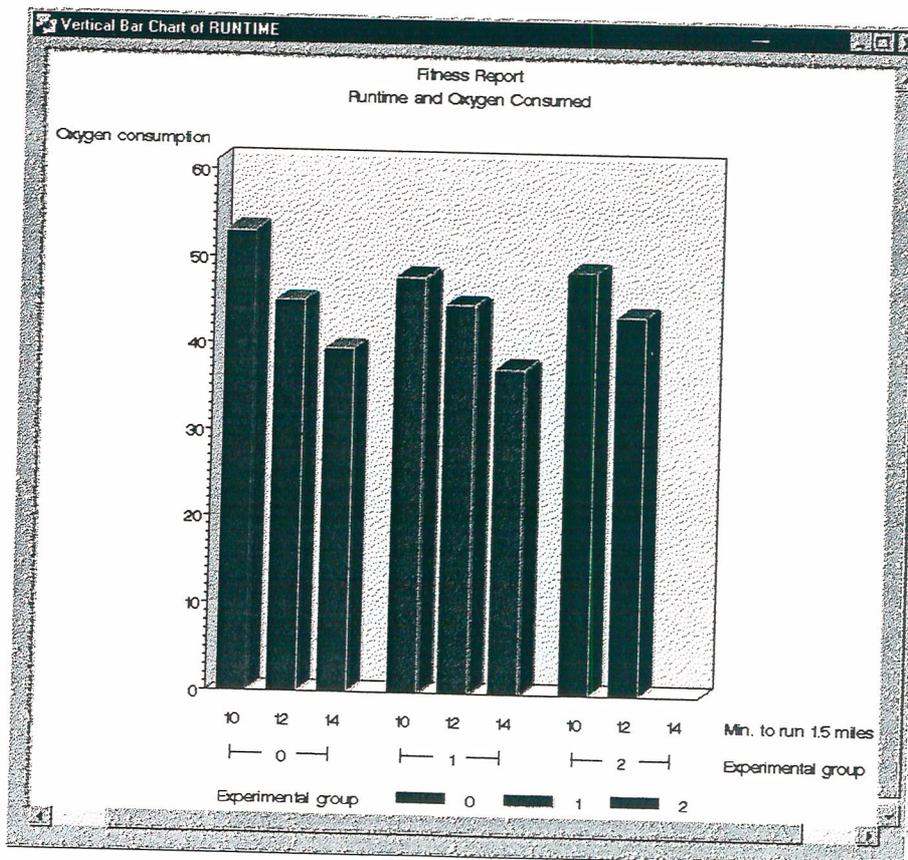


Figura 8.16 Gráfico de Barras Verticais

Exercício 8:

Construa um gráfico de barras verticais da quantidade de chuvas nos meses a partir dos dados do arquivo **chuva.sas**.

8.2 Diálogo Gráfico de Setores

Um gráfico de setores exibe seus dados na forma de um disco bidimensional ou tridimensional, dividido em fatias. O tamanho de cada fatia indica a contribuição relativa de cada parte para o todo. Selecione **Pie Chart** do menu **Graphs** para exibir seus dados em gráfico de setores.

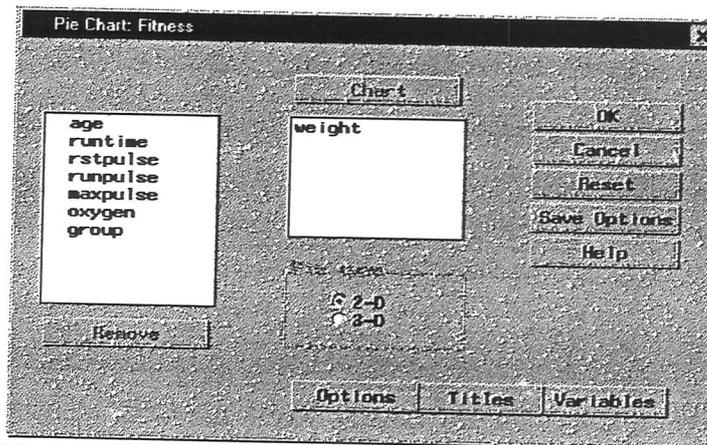


Figura 8.17 Diálogo Gráfico de Setores

Lista de Candidatos

A lista de candidatos mostra todas as variáveis do conjunto de dados. Você pode selecionar variáveis da lista para exibi-las em um gráfico de setores. Note que selecionando variáveis não contíguas da lista de candidatos, somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais, você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão Chart

Selecione as variáveis da lista de candidatos e clique no botão **Chart** para produzir um gráfico de setores para cada variável.

Tipo de Setor

Selecione **2-D** ou **3-D** sobre o **Pie Type (Tipo de Setor)** para especificar se você quer exibir gráficos bidimensionais ou tridimensionais.

Opções do Gráfico de Setores

As seguintes opções estão disponíveis para ajustar o seu gráfico de setores.

- **Options:** Controla a aparência do gráfico de setores.

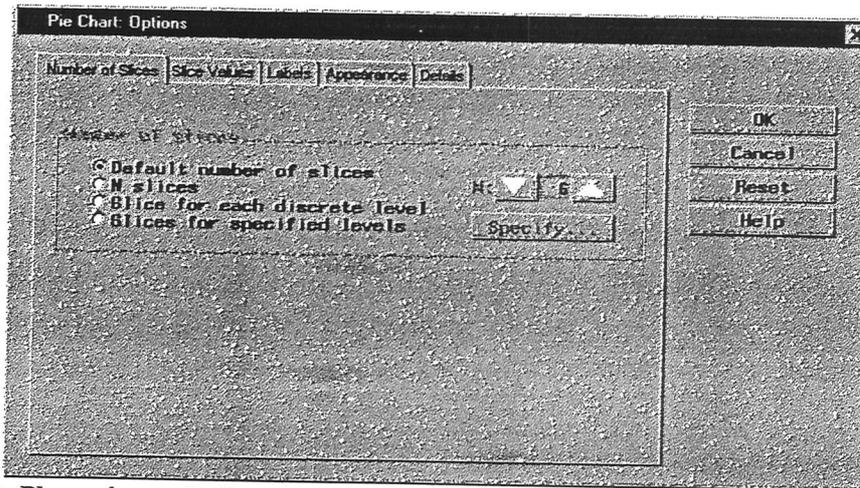


Figura 8.18 Placa número de Fatias

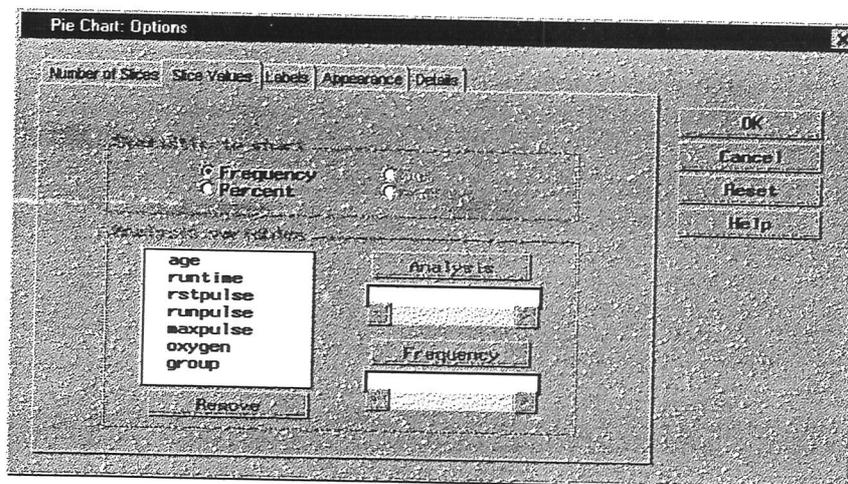


Figura 8.19 Placa Valores das Fatias

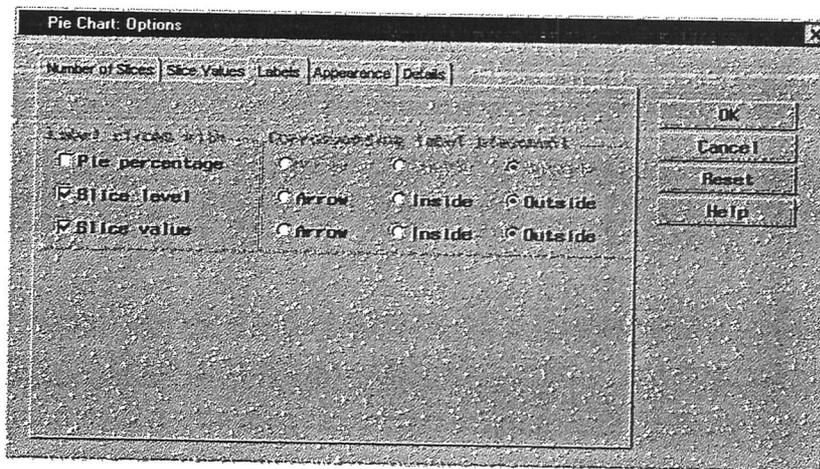


Figura 8.20 Placa Rótulos

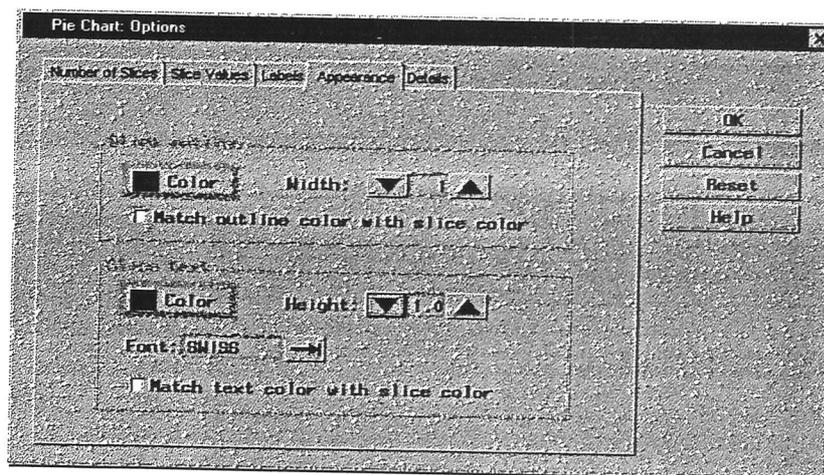


Figura 8.21 Placa Aparência

- Títulos: Especifica os títulos dos resultados.
- Variáveis: Especifica POR grupos de variáveis.

8.2.1. Exemplo:

Neste exemplo, você cria um gráfico de setores de um conjunto de dados Fitness, o mesmo usado no exemplo anterior.

Especificar uma variável para o Gráfico de Setores.

Para especificar uma variável que vai ser representada e o tipo do gráfico, siga as seguintes etapas:

1. Selecione **Graphs Pie chart**.

2. Selecione runtime da lista de candidatos e clique em **Chart** para fazer minutos para corre 1.5 milhas a variável a ser representada pelo gráfico.
3. Selecione **3-D** sob **Pie type** para especificar um gráfico tridimensional.

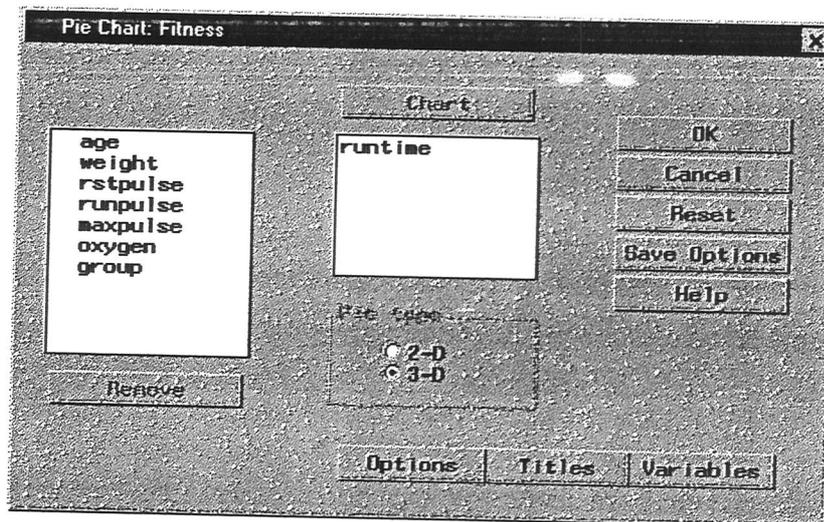


Figura 8.22 Variáveis e Tipo do Gráfico de Setores

Especificar Opções do Gráfico de Setores

Para especificar suas opções de gráfico de setores, assim como o número de fatias, siga as etapas seguintes:

1. Clique no botão **Options** para exibir o diálogo de opções do Gráfico de Setores.
2. Na placa **Number of Slices**, projete um gráfico com dez fatias, selecionando **N slices** e clicando na seta para cima até o número 10 ficar visível.

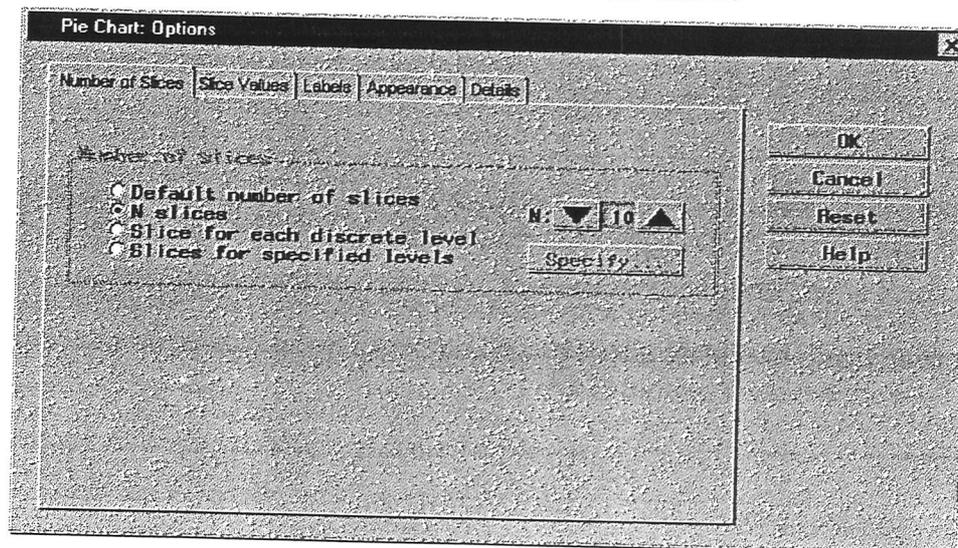


Figura 8.23 Número de Fatias do Gráfico de Setores

3. Na placa **Slice Values**, selecione **Percent** sob **Statistic to chart** em ordem para representar a porcentagem de cada tempo de corrida na relação para cada variável.

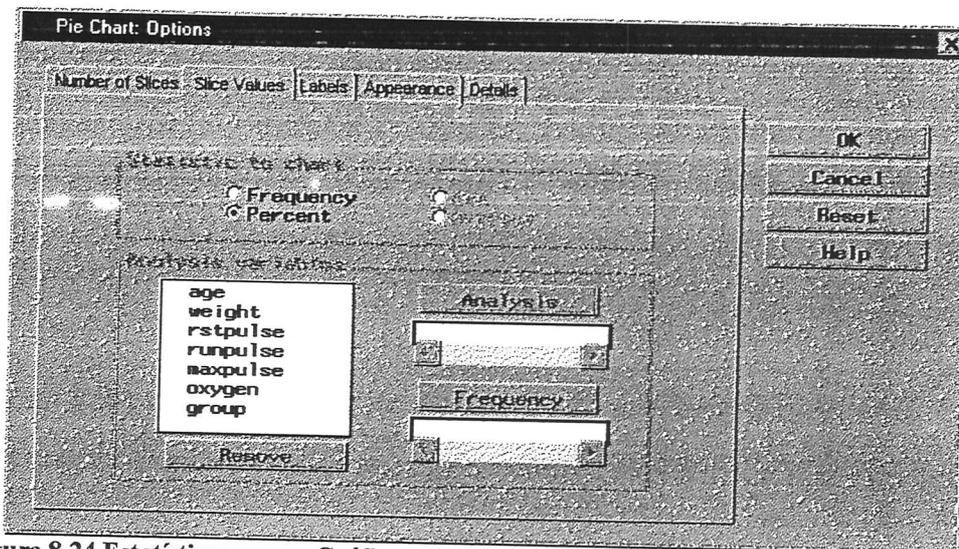


Figura 8.24 Estatísticas para o Gráfico

4. Na placa **Labels**, selecione **Slice level** sobre **Labels slices with**. Selecione **Arrow** sob **Corresponding label placement**. Cada fatia indica um tempo de corrida, e cada rótulo é colocado fora do disco, com uma seta apontando para a fatia correspondente.

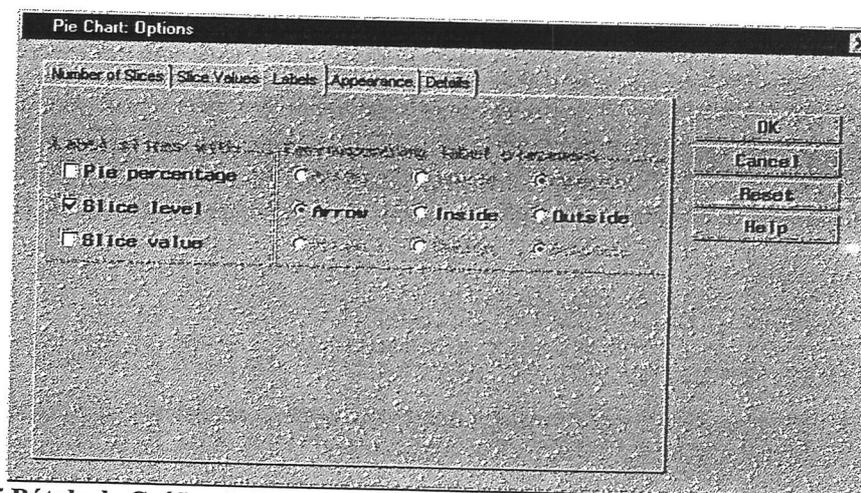


Figura 8.25 Rótulo do Gráfico de Setores

5. Na placa **Details**, desmarque **Show default heading above chart** sob **Chart options**. Você fornece um novo título no diálogo **Titles**.

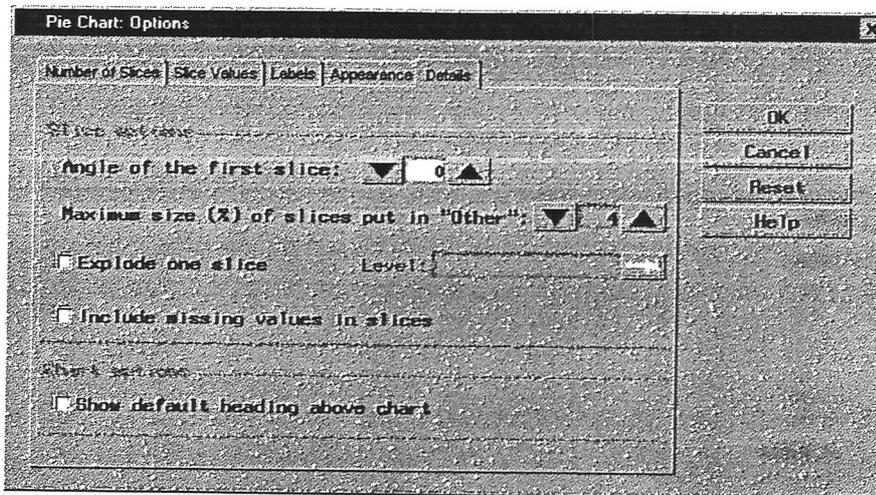


Figura 8.26 Desmarcando o Título Padrão

6. Clique em **OK** para salvar suas mudanças e retorne para o diálogo do Gráfico de Setores.

Especifique os Títulos do Gráfico de Setores

Para especificar os títulos para o seu gráfico de setores, siga as seguintes etapas:

1. Clique no botão **Titles** no diálogo do gráfico de setores.
2. Na placa **pie chart**, escreva **Percentage of Each runtime** (Percentual de Cada Tempo de Corrida) no primeiro campo.

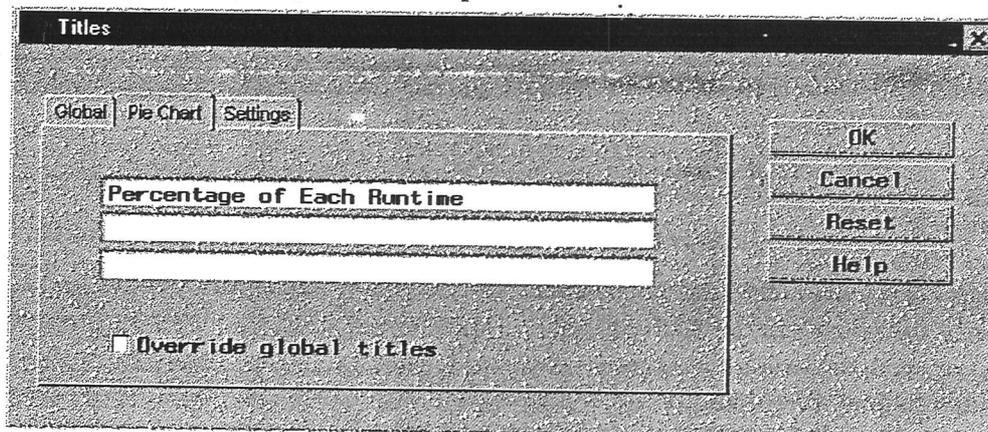


Figura 8.27 Título do Gráfico de Setores

3. Clique em **OK** para salvar suas mudanças no título.

Gerando o Gráfico de Setores

Para exibir seu gráfico de setores, clique em **OK** no diálogo Pie Chart.

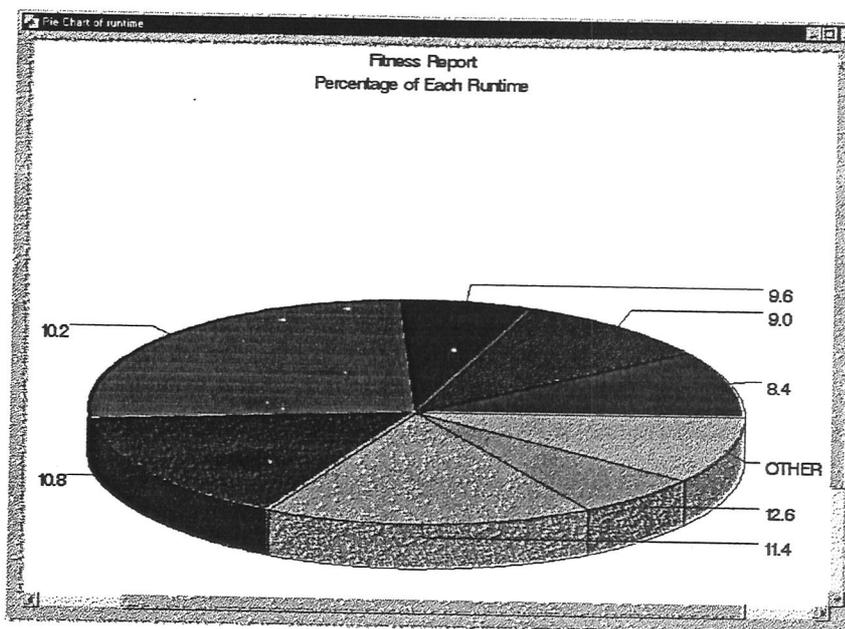


Figura 8.28 Gráfico de Setores 3-D

Exercício 9:

Construa um gráfico de setores com os dados do arquivo **composicao.xls**.

8.3 Diálogo Histograma

Selecione **Histogram** do menu **Graphs** para produzir um histograma dos dados da sua tabela corrente. Histogramas permitem que você explore seus dados através da exibição da distribuição de uma variável particular sobre vários intervalos ou classes. Assim, você pode ganhar uma indicação da forma da distribuição e se os dados estão distribuídos simetricamente. Um histograma comparativo é produzido se você especificar a classificação da variável.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las no **histograma**. Um "C" que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla **Ctrl** enquanto clica nas variáveis.

Botão Remove

O botão **remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão Analysis

Selecione uma variável da lista de candidatos e clique no botão **Analysis** para adicioná-la na lista de variáveis em que será produzido o histograma.

Botão Class

Para definir subgrupos, selecione uma variável de classificação e clique no botão **Class**.

Opções de Histograma

As seguintes opções são avaliadas para ajustar a sua análise.

- **Method**: Escolhe a forma da variância.
- **Display**: Especifica a opção para a aparência do histograma.
- **Fit**: Especifica o ajuste da distribuição.
- **Titles**: Especifica os títulos para a análise.
- **Variables**: Especifica variáveis POR grupo e frequência.

8.4 Diálogo Box Plot

Selecione **Box Plot** no menu **Graphs** para produzir um box plot dos dados da tabela corrente. Os Box Plots são uma técnica para exibir dados unidimensionais e resumo de suas características. Uma caixa-e-linhas exibem os dados através de: a mediana é representada pela linha horizontal dentro da caixa; o topo e a base da caixa representa o 3º quartil (75%) e o 1º quartil (25%), respectivamente. A distância entre estes dois é a amplitude interquartílica (IRQ).

No padrão do estilo de construção linhas são desenhadas da margem superior da caixa até a observação máxima dentro da grade inferior. A grade superior está localizada 1.5 vezes da IQR acima do 3º quartil, e a grade inferior está 1.5 vezes abaixo do 1º quartil. As grades não são exibidas. Observações que caem além das grades são identificados individualmente como símbolos.

Se você especificar uma variável de classificação simples, um gráfico de caixa será criado com uma caixa-e-linhas para cada nível das variáveis de classificação.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las no **Box-Plot**. Um "C" que precede a variável indica que ela é uma variável não numérica (caracter). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão Analysis

Selecione uma variável da lista de candidatos e clique no botão **Analysis** para criar um box plot.

Botão Class

Selecione uma variável da lista de candidatos e clique no botão **Class** para selecionar uma variável classificatória. Cada nível dessas variáveis classificatórias é representado como uma caixa separada no box plot.

Opções de Box Plot

As seguintes opções são avaliadas para caracterizar a sua análise.

- **Display:** Especifica a opção para a aparência do box plot.
- **Titles:** Especifica os títulos que aparecem nos resultados.
- **Variables:** Especifica variáveis POR grupo.

8.5 Diálogo Gráfico de Probabilidade

Selecione **Probability Plot** do menu **Graphs** para produzir o gráfico de probabilidade da tabela de dados corrente. O gráfico de Probabilidade compara valores ordenados da variável dentro dos percentis da distribuição teórica especificada. Se a distribuição dos dados adapta-se a distribuição teórica, os pontos no gráfico formarão um padrão linear. As distribuições avaliadas são a Normal, a Lognormal, Exponencial e Weibull. Você pode usar o diálogo **Parameters** para especificar os valores dos parâmetros associados com cada distribuição, de outra forma, os parâmetros necessários serão estimados. A linha de referência correspondente para especificar a distribuição teórica é produzindo o gráfico.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las no **Gráfico de Probabilidade**. Um “C” que procede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão Analysis

Selecione uma variável para criar um gráfico de probabilidade e clique no botão **Analysis**.

Distribuição

Selecione a distribuição teórica.

Opção de Probabilidade

As seguintes opções são válidas para personalizar seu gráfico de probabilidade.

- Method: Escolhe a forma da variância.
- Display: Especifica a cor e opções de eixo.
- Parameters: Especifica os valores dos parâmetros para a distribuição teórica.
- Titles: Especifica os títulos para a análise.
- Variabels: Especifica as variáveis POR grupo e frequência.

8.6 Diálogo Gráfico de Dispersão

Selecione **Scatter Plot** do menu **Graphs**, então selecione uma das seguintes opções:

- **Two-Dimensional**: para produzir um gráfico de dispersão bidimensional para os dados da tabela corrente.
- **Three-Dimensional**: para produzir um gráfico de dispersão tridimensional para os dados da tabela corrente.

Se você especificar mais que uma variável para qualquer eixo, um gráfico é produzido para cada combinação das variáveis.

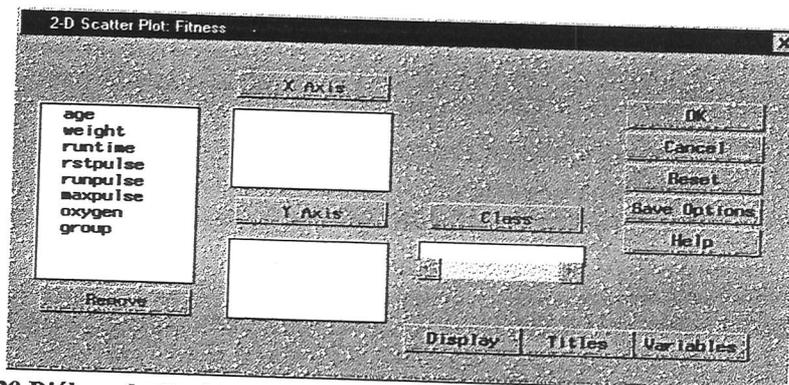


Figura 8.29 Diálogo do Gráfico de dispersão 2-D

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las no **Gráfico de dispersão**. Um “C” que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **Remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão X Axis

Você deve especificar uma ou mais variáveis para o eixo X.

Botão Y Axis

Você deve especificar uma ou mais variáveis para o eixo Y.

Botão Z Axis (Gráficos de Dispersão Tridimensionais)

Para gráficos tridimensionais, você deve especificar uma ou mais variáveis para o eixo Z.

Botão Class (Gráficos de Dispersão Bidimensionais)

Especifica uma variável classificatória para definir subgrupos. Cada nível da variável classificatória é representado por um diferente símbolo no gráfico de dispersão.

Opções de Gráficos de Dispersão

As seguintes opções estão disponíveis para personalizar suas análises.

- Display (2D Plots): Especifica as opções de cor e eixos para um gráfico de dispersão bidimensional.

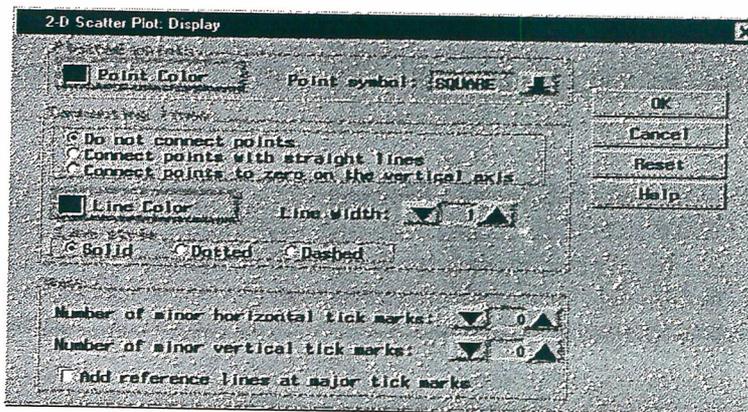


Figura 8.30 Gráfico de Dispersão 2-D: Diálogo de Exibição

- Display (3D Plots): Especifica as opções de cor e eixos para um gráfico de dispersão tridimensional.

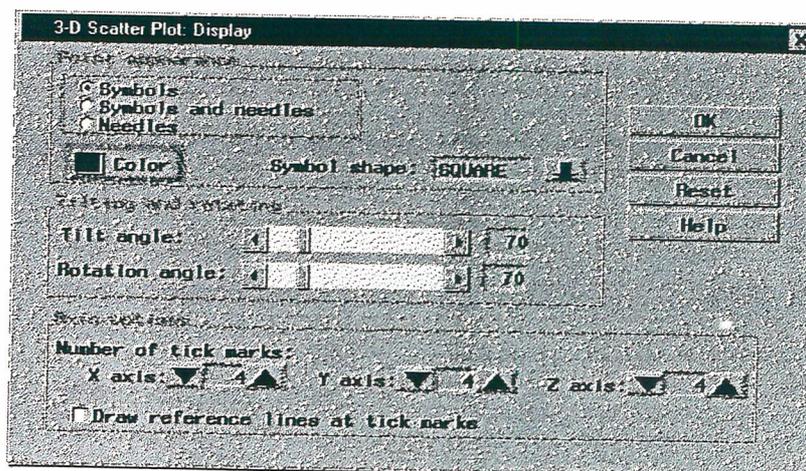


Figura 8.31 Gráfico de Dispersão 3-D: Diálogo de Exibição

- **Titles:** Especifica os títulos para a análise.

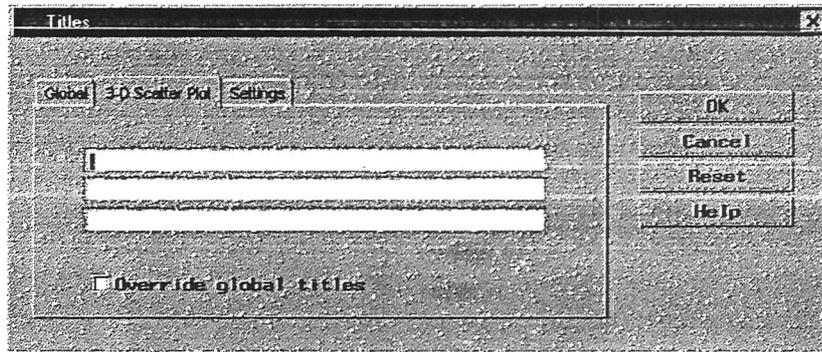


Figura 8.32 Diálogo Títulos: Placa do Gráfico de Dispersão 3-D

- **Variables:** Especifica POR grupo de variáveis

8.6.1 Exemplo:

Neste exemplo, você usa o conjunto de dados de Aptidão Física (Fitness) como base de seu gráfico de dispersão.

Especifique as Variáveis do Gráfico de Dispersão

Para especificar as variáveis a serem esboçadas, siga as seguintes etapas:

1. Selecione **Graphs Scatter plot Two-Dimensional**.
2. Selecione *age* da lista de candidatos e clique no **X Axis** para fazer a idade em anos a variável do eixo x.
3. Selecione *runtime* da lista de candidatos, e clique no **Y Axis** para fazer para fazer os minutos para correr 1.5 milhas a variável do eixo y.

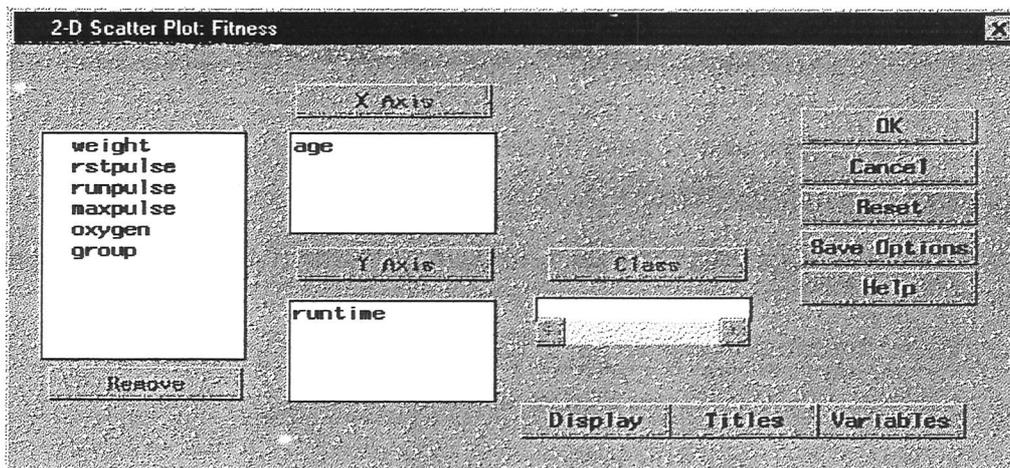


Figura 8.33 Variáveis do Gráfico de Dispersão

Especificar as Opções do Gráfico de Dispersão

Para especificar suas opções de gráfico de dispersão, siga as seguintes etapas:

1. Clique no botão **Display** para exibir o diálogo do gráfico de dispersão.

2. Sob **Plotted points**, clique no botão **Point Color**. Selecione **Red** da lista de cores para plotar pontos vermelhos no seu gráfico de dispersão. Clique em **OK**.
3. Clique na seta para baixo próxima do **Point symbol** e selecione **DOT** da lista. Isto faz o seu gráfico de dispersão se exibido como pontos.
4. Sob **Axes**, selecione **Add reference lines at major tick marks**. Isto exhibe uma grade no seu gráfico que você pode orientar os pontos nos eixos.

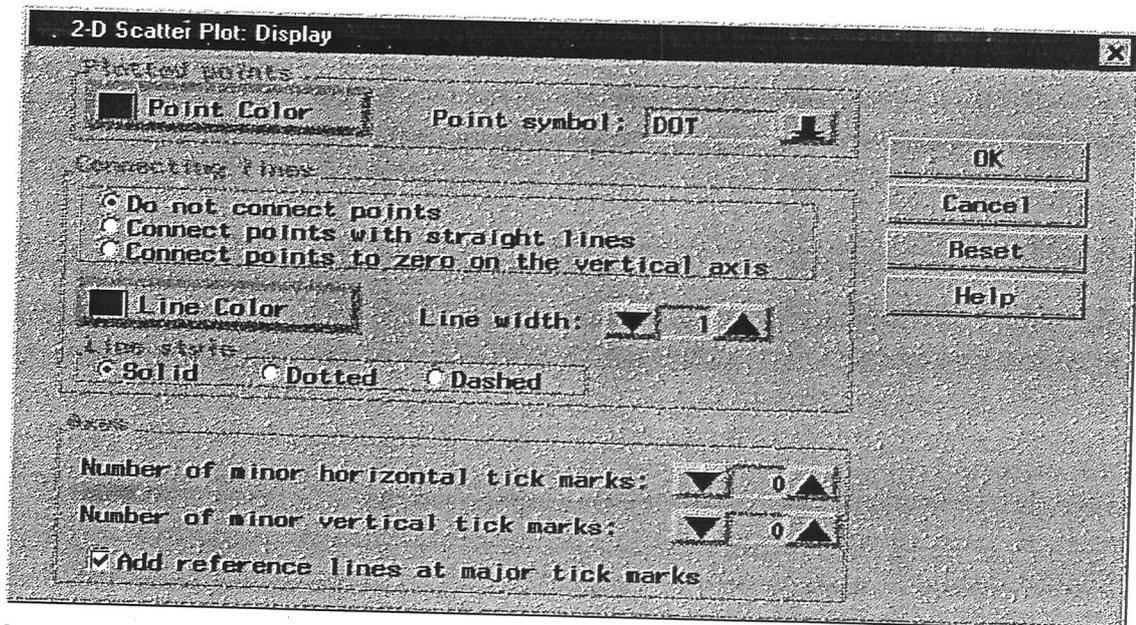


Figura 8.34 Display Options

5. Clique em **OK** para salvar suas mudanças de exibição.

Especificar os Títulos do Gráfico de Dispersão

Para especificar os títulos do seu gráfico de dispersão, siga as seguintes etapas:

1. Clique no botão **Titles** no diálogo de Gráfico de Dispersão.
2. Na placa **Scatter Plot**, escreva **Age versus Runtime** no primeiro campo.

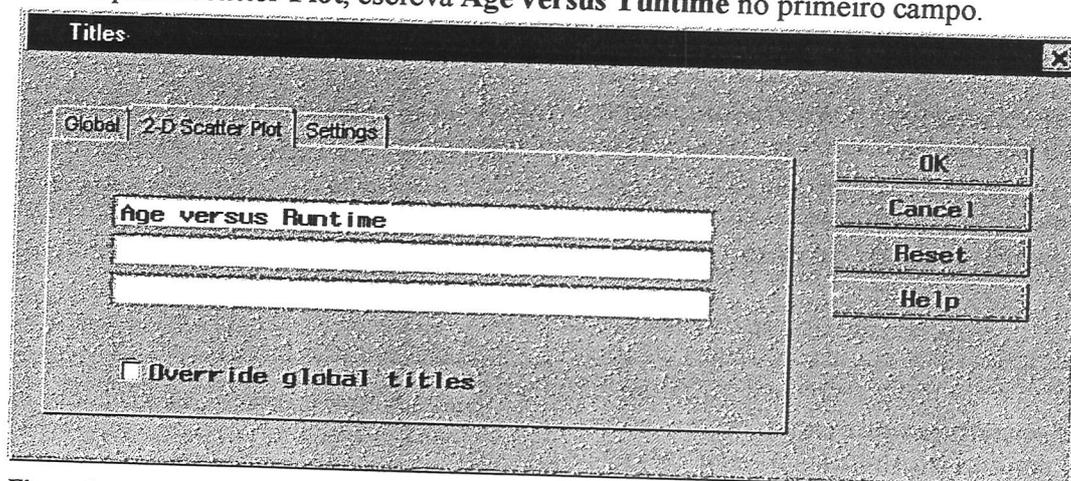


Figura 8.35 Título do Gráfico de Dispersão

3. Clique em **OK** para salvar suas mudanças no título.

Gerando o Gráfico de Dispersão

Para exibir o gráfico de dispersão, clique em **OK** no diálogo do Gráfico de Dispersão.

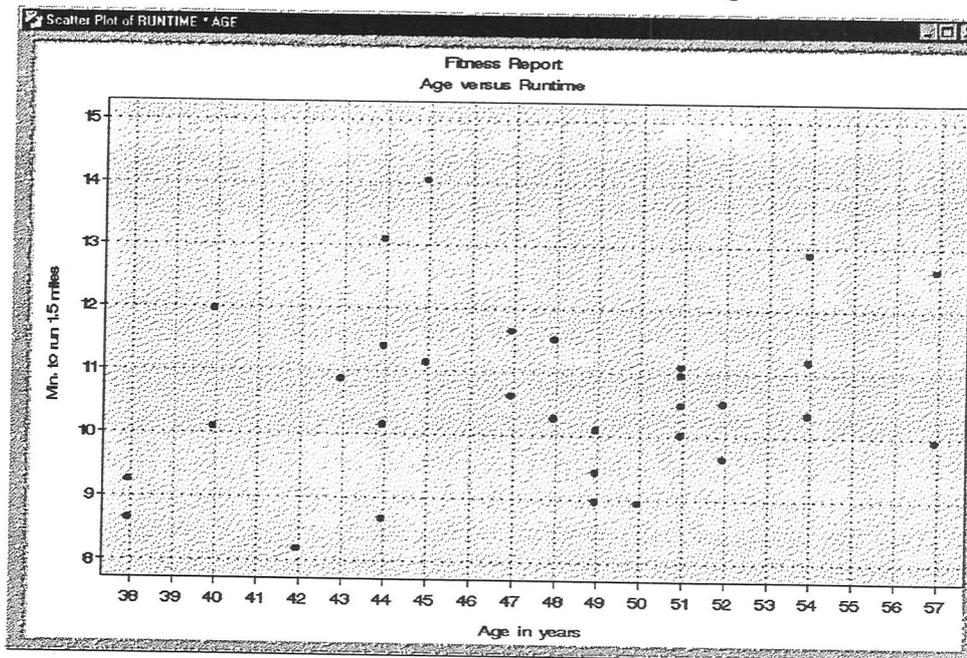


Figura 8.36 Gráfico de Dipersão 2-D.

Exercício 10:

Construa um gráfico de dispersão das variáveis MSR e MST do arquivo **mandioca.xls**.

8.7 Diálogo Gráfico de Contorno (Contour Plot)

Selecione **Countour Plot** do menu **Graphs** para produzir um gráfico de contorno para a tabela corrente. Gráficos de contornos permitem que você represente tridimensionalmente em duas dimensões. Linhas ou áreas em um gráfico de contorno representam os níveis de magnitude para a variável Z correspondente para a posição (X,Y) no plano. Você pode usar o gráfico de contorno para analisar tendências em seus dados quando eles contêm muitos picos e vales para serem observados com um gráfico de superfície. Você pode também examinar dados em que os níveis, não a forma, dos dados são importantes.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las no **Gráfico de Contornos**. Um "C" que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer

selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **Remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão X Axis

Especifica variável para o eixo X.

Botão Y Axis

Especifica variável para o eixo Y.

Botão Z Axis

Especifica uma ou mais variáveis respostas para o eixo Z. Um gráfico separado é produzido para cada variável que você especificar.

Opções de Gráficos de Contorno

Os seguintes diálogos estão disponíveis para personalizar seu gráfico de contorno:

- Display: Especifica várias opções no gráfico de contorno.
- Interpolate: Especifica como a grade de (X,Y) dos dados é criada.
- Titles: Especifica os títulos para a análise.
- Variables: Especifica POR grupo de variáveis

8.8 Diálogo Gráfico de Superfície

Selecione **Surface Plot** do menu **Graphs** para produzir um gráfico de superfície para os seus dados. Você pode usar o diálogo **Surface Plot** para produzir gráfico de superfície tridimensional. Para gráfico de superfícies os valores X e Y deveriam formar uma grade igualmente espaçada para valores horizontais. Se as variáveis que você especificou para o eixo X e eixo Y não estão na forma, os dados são processados para formar uma grade. Para pontos da grade em que não existe um valor Z, o necessário valor é interpolado. Você pode usar o diálogo **Interpolate** para escolher um método de interpolação. Um gráfico separado é produzido para cada variável Z especificada.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibí-las no **Gráfico de Superfície**. Um "C" que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **Remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão X Axis

Especifica variável para o eixo X.

Botão Y Axis

Especifica variável para o eixo Y.

Botão Z Axis

Especifica uma ou mais variáveis respostas para o eixo Z.

Opções do Gráfico de Superfície

Os seguintes diálogos estão disponíveis para personalizar seu gráfico de superfície:

- Display: Especifica opções de cor e eixo.
- Interpolate: Escolhe um método de interpolação.
- Titles: Especifica os títulos para a análise.
- Variables: Especifica POR grupo de variáveis.

9. Tarefas Estatísticas

9.1 Estatística Descritiva

9.1.1 Introdução

Estatísticas descritivas e gráficos são muitas vezes utilizados na fase inicial de uma análise estatística. Estas ferramentas são úteis para identificar relações nos dados e para determinar relações para análises futuras.

O Analyst fornece amplo espectro de estatísticas descritivas e de exibições gráficas. O menu para Estatística Descritiva aparece na figura 9.1.1

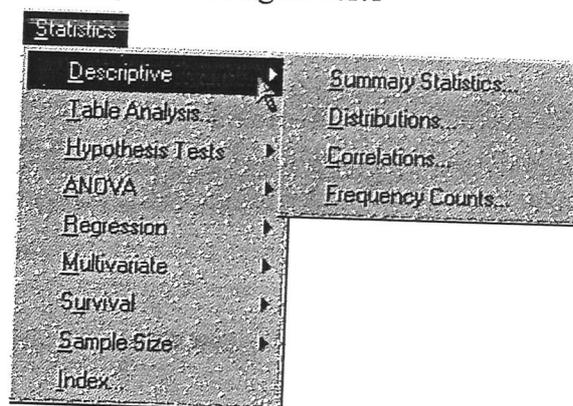


Figura 9.1.1 Menu para Estatística Descritiva

9.1.2 Cálculo de Frequências

Selecione **Descriptive** do menu **Statistics**, então selecione **Frequency Counts** para mostrar o diálogo **Frequency Counts**. Você pode computar tabelas de frequência em classificação simples (one-way) para uma ou mais das variáveis no conjunto de dados. Para cada valor da variável, a frequência é produzida. A porcentagem, a frequência acumulada, e a porcentagem acumulada são produzidas opcionalmente.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibir na análise. Um "C" que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla **Ctrl** enquanto clica nas variáveis.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto a lista de candidatos, na qual todas as variáveis estão no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista por clique duplo nelas.

Botão Frequencies

Especifica as variáveis para as quais você gostaria de computar frequências. Uma tabela é produzida para cada variável que você especifica.

Opções para Calcular Frequência

Os seguintes diálogos estão disponíveis para personalizar sua análise:

- ◆ **Input:** especifica a ordem para ajustar sua análise.
- ◆ **Plots:** produz diagrama de barras.
- ◆ **Tables:** especifica informações para incluir na tabela de frequência.
- ◆ **Titles:** especifica os títulos que aparecem nos resultados.
- ◆ **Variables:** especifica por grupo e variáveis de contagem.

Exemplo 9.1.1

Os dados considerados foram retirados do Abstract Estatístico dos Estados Unidos de 1995 e referem-se as taxas de natalidade e mortalidade infantil nos Estados Unidos em 1992. A informação é proveniente dos 50 estados e do Distrito de Columbia. Os estados são agrupados por região.

Abrir o arquivo de dados

Os dados fazem parte da Livraria de exemplos do Analyst. Para acessar o conjunto de dados segue-se as seguintes etapas:

1. Selecione **Tools Sample Data ...**
2. Selecione Bthdth92.
3. Clique **OK** para criar o conjunto de dados no seu diretório Sasuser
4. Selecione **File Open By SAS Name ...**
5. Selecione Sasuser da lista de **Libraries**.
6. Selecione Bthdth92 da lista de membros
7. Clique **OK** para abrir o conjunto de dados na sua tabela de dados

Solicitar Contagem de Frequências (*Request Frequency Count*)

Para solicitar Contagem de Frequências, segue-se as seguintes etapas:

1. Selecione **Statistics Descriptive Frequency Counts...**
2. Selecione região como a variável da lista de candidatos para a contagem de frequências.

A análise por definição fornece a informação desejada. Note que pode-se usar o botão Input para selecionar a ordem específica pela qual os valores das variáveis são listados.

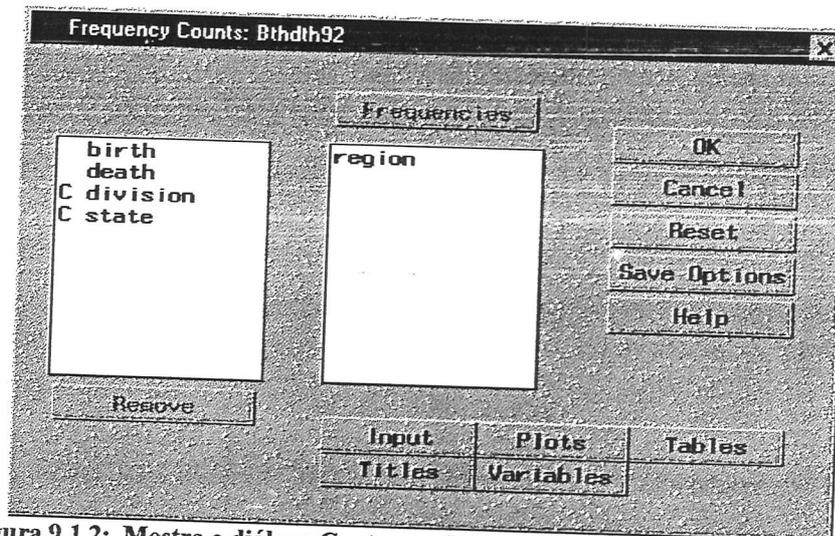


Figura 9.1.2: Mostra o diálogo Contagem de Frequência com região como a variável para a frequência especificada.

Solicitar Diagrama Horizontal de Barras (*Request a Horizontal Bar Chart*)

Para gerar um Diagrama Horizontal de Barras em acréscimo à Contagem de Frequências, segue-se as seguintes etapas:

1. Selecione o botão **Plots**.
2. Selecione **Horizontal**, como mostra a figura 9.1.3.
3. Clique **OK** para fechar a caixa **Plots**.

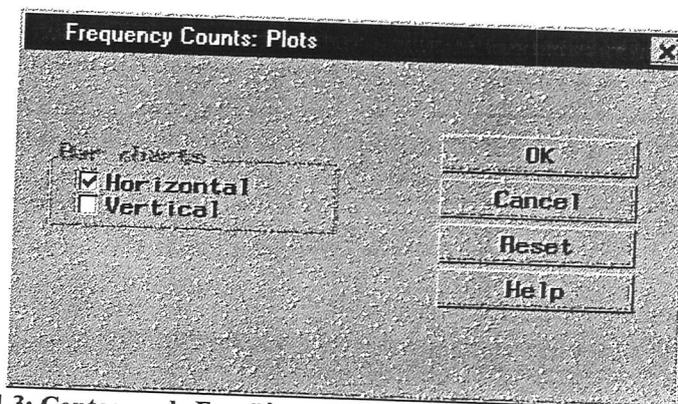


Figura 9.1.3: Contagem de Frequências: Caixa Plots.

4. Clique **OK** na Caixa Principal Frequency Counts para executar a análise.

Análise dos Resultados (*Review the Results*)

Os resultados são apresentados na árvore de projeto abaixo da pasta **Frequency Counts**, como mostra a figura 9.1.4. Os três nós representam o resultado da contagem de frequências, o diagrama horizontal de barras e a instrução de programação do SAS (intitulado **Code**), que gerou o resultado.

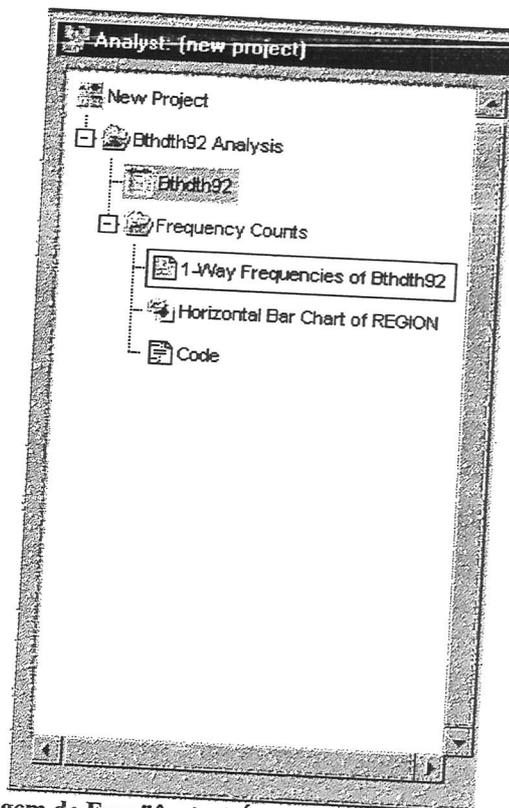


Figura 9.1.4: Contagem de Freqüências: Árvore de Projeto.

Pode-se dar um duplo clique em qualquer nó da árvore de projeto para ver o seu conteúdo numa janela separada. Note que o primeiro resultado gerado é exibido sempre.

A Figura 9.1.5 mostra a tabela de contagem de freqüências para as variável região.

The FREQ Procedure				
region	Frequency	Percent	Cumulative Frequency	Cumulative Percent
MW	12	23.53		
NE	9	17.65	12	23.53
S	17	33.33	21	41.18
W	13	25.49	38	74.51
			51	100.00

Figura 9.1.5: Contagem de Freqüências: Freqüências em classificação Simples (One- Way) da variável região.

A tabela mostra que aproximadamente 33% das observações do grupo de dados estão localizados na região Sul; e 25% das observações estão localizadas nas regiões Oeste

e Meio-Oeste, respectivamente. Aproximadamente 18% das observações estão localizadas na região Nordeste.

Para mostrar o diagrama de barras da contagem de frequências, dê um duplo clique no nó **Horizontal Bar Chart** de **REGION** (Figura 9.1.6)

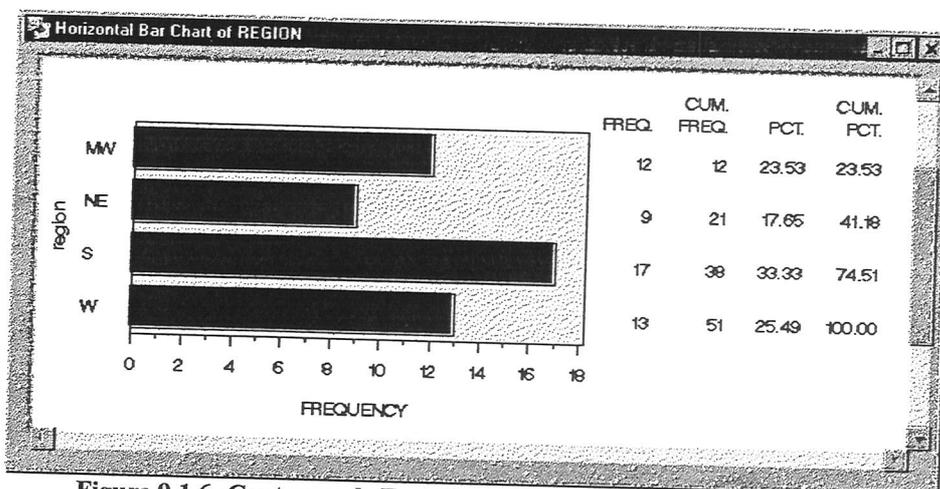


Figura 9.1.6: Contagem de Frequências: Diagrama de Barras Horizontal de Região

Exercício 11.

Calcule frequência das variáveis do arquivo **pesovelo.xls**.

9.1.3 Resumo Estatístico

Selecione **Descriptive** do menu **Statistics**, então selecione **Summary Statistics** para produzir um resumo estatístico dos dados na tabela atual. No diálogo **Summary Statistics** você pode produzir informações resumidas para variáveis quantitativas. Na lista de candidatos, selecione as variáveis a serem analisadas e clique no botão **Analysis**. Você também pode especificar uma classificação de variáveis para definir subgrupos.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibir na análise. Um "C" que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla **Ctrl** enquanto clica nas variáveis.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto a lista de candidatos, na qual todas as variáveis estão no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista por clique duplo nelas.

Botão Analysis

Selecione as variáveis quantitativas a serem analisadas e clique no botão **Analysis**.

Botão Class

Para definir subgrupos, selecione uma variável de classificação e clique no botão **Class**.

Opções de Resumo Estatístico

As opções seguintes estão disponíveis para personalizar sua análise.

- **Statistics**: especifica quais estatísticas de resumo computar.
- **Plots**: especifica um histograma ou um box-and-whisker.
- **Output**: controla a aparência da saída de resultados e do gráfico de caixa e linhas (box-plot).
- **Save Data**: salva estatísticas para um conjunto de dados.
- **Titles**: especifica os títulos que aparecem no resultado.
- **Variables**: especifica peso, por grupo e frequência de variáveis.

Exemplo 9.1.2

Consideremos os dados do exemplo 9.1.1. Na tarefa resumo estatístico (tais como a média, o desvio padrão, os valores mínimos e máximos) são desejados para as taxas de natalidade e mortalidade infantil para cada região. Em adição os box-and-whisker plots são solicitados.

Solicitar Resumo Estatístico (*Request Summary Statistics*)

Para solicitar Resumo Estatístico, siga os seguintes passos:

1. Selecione **Statistics Descriptive** → **Summary Statistics**.
2. Selecione as variáveis de análise **birth** e **death** da lista de candidatos.

Pode-se especificar uma variável de classificação para definir grupos dentro dos dados. Quando você especifica uma variável de classificação, o Analyst gera um resumo estatístico para a variável de análise para cada nível da variável de classificação. Selecione região como a variável de classificação.

A figura 9.1.7 mostra a caixa de diálogo Summary Statistics com birth e death especificados como a variável de análise e região especificada como a variável de classificação.

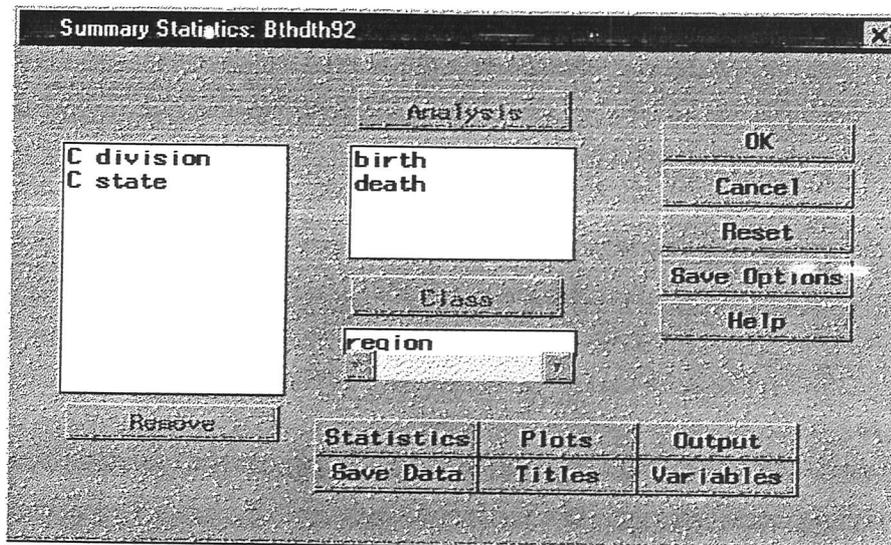


Figura 9.1.7: Diálogo Summary Statistics.

Solicitar Gráficos Box-and-Whisker

Para solicitar box-and-whisker plots, siga os seguintes passos:

1. Clique o botão **Plots**.
2. Selecione **Box-&-whisker plot**.
3. Clique **OK**.

A Figura 9.1.8 mostra a caixa de diálogo Plots com **Box-&-whisker plot** selecionado.

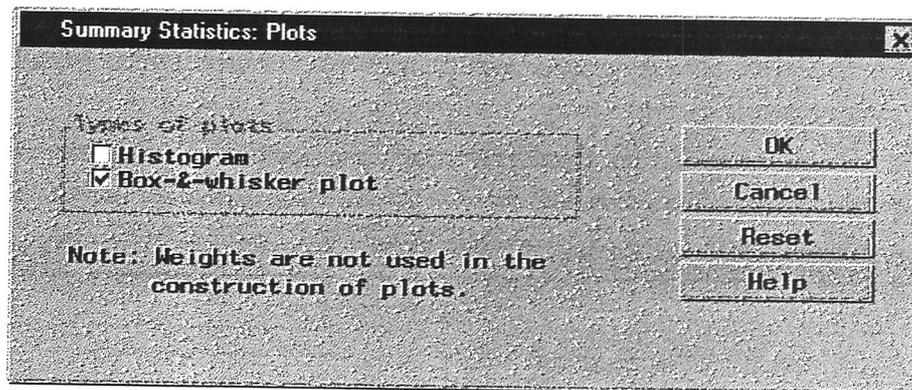


Figura 9.1.8: Resumo Estatístico: Diálogo Plots.

Para executar a análise, clique **OK** na caixa principal.

Exercício 12.

Calcule algumas estatísticas descritivas do banco de dados **pesovelo.xls**.

Análise dos Resultados (*Review the Results*)

Os resultados são apresentados na árvore de projetos abaixo da pasta Resumos Estatísticos (**Summary Statistics**), como mostra a figura 9.1.9. Os quatro ícones representam o resultado dos resumos estatísticos, o box-and-whisker para cada variável de análise, e a instrução de programação do SAS (intitulado **Code**), que gerou o resultado.

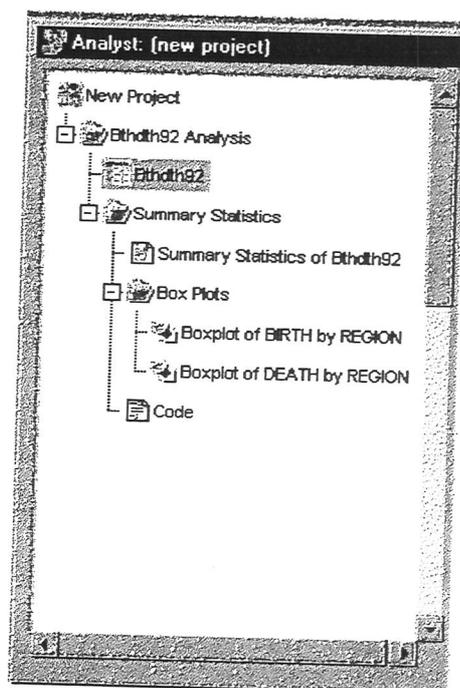


Figura 9.1.9: Resumo Estatístico: Árvore de Projeto.

Um duplo clique em qualquer dos ícones mostra a informação correspondente em uma janela separada.

A figura 9.1.10 mostra, para cada valor da variável de classificação região, o número de observações, a mediana, o desvio padrão, e os valores mínimo e máximo de cada variável de análise. A região Oeste tem a mais elevada taxa de natalidade (16.89) e a região Sul tem a mais elevada taxa de mortalidade (10.15).

The MEANS Procedure							
region	N	Variable	N	Mean	Std Dev	Minimum	Maximum
MW	12	birth	12	14.8250000	0.7581377	13.7000000	16.5000000
		death	12	8.5916667	1.0974833	7.1000000	10.2000000
NE	9	birth	9	14.3666667	0.8930286	13.0000000	15.9000000
		death	9	7.3777778	1.2194033	5.6000000	9.0000000
S	17	birth	17	15.4647059	1.4924565	12.3000000	18.7000000
		death	17	10.1529412	2.6241946	7.8000000	19.6000000
W	13	birth	13	16.8923077	2.1864970	14.0000000	20.5000000
		death	13	7.4769231	0.9670866	5.9000000	8.9000000

Figura 9.1.10: Resumos Estatísticos: Estatísticas de natalidade e morte.

A Figura 9.1.11 mostra o box-and-whisker plot para a variável natalidade para cada nível da variável região.

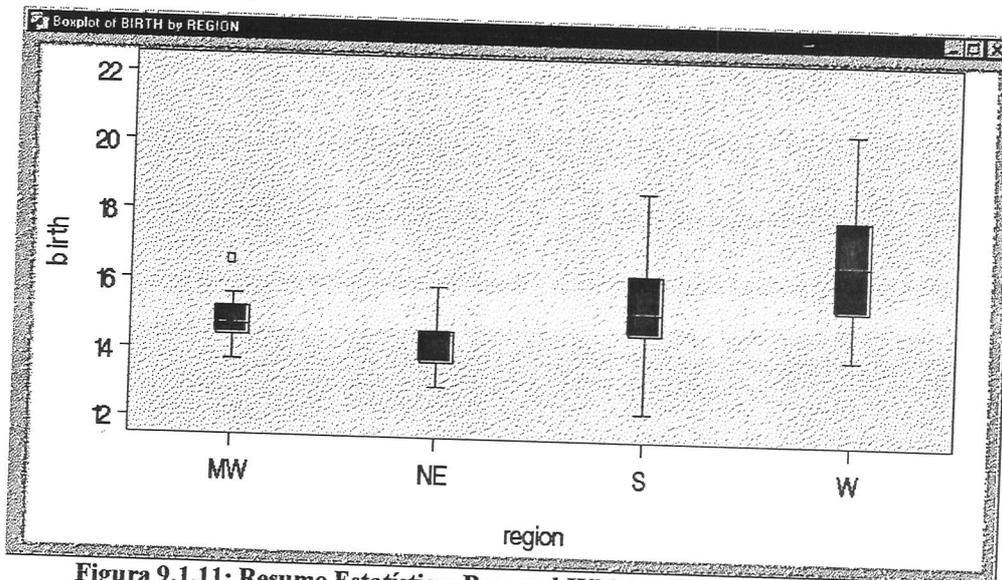


Figura 9.1.11: Resumo Estatístico: Box-and-Whisker Plot para Taxa de Natalidade por Região.

Este plot revela um possível outlier na taxa de natalidade para a região Meio-Oeste (region='MW'). A região Oeste (region='W') é notadamente a região com a taxa de natalidade mais elevada.

9.1.4 Distribuições

Selecione **Descriptive** do menu **Statistics**, então selecione **Distributions** para produzir distribuições estatísticas dos dados na tabela atual. O diálogo **Distributions** permite que você examine a distribuição de uma ou mais variáveis no conjunto de dados. A análise mais simples produz estatísticas descritivas para cada variável selecionada, inclusive momentos e quantis. Você também pode especificar um variável de classificação para definir subgrupos.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibir na análise. Um "C" que precede a variável indica que ela é uma variável não numérica (caracter). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto a lista de candidatos, na qual todas as variáveis estão no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista por clique duplo nelas.

Botão Analysis

A lista Analysis contém as variáveis que serão analisadas. Uma análise separada é executada para cada variável na lista de análise. Selecione uma variável da lista de candidatos e clique no botão **Analysis** para colocá-la na lista Analysis.

Botão Class

Para definir subgrupos, selecione uma variável de classificação e clique no botão **Class**.

Opções de Distribuições

As seguintes opções estão disponíveis para personalizar sua análise.

- ◆ **Method**: especifica o divisor de variância a ser utilizado.
- ◆ **Plots**: produz uma variedade de plots.
- ◆ **Fit**: ajusta os dados para distribuições específicas.
- ◆ **Save Data**: salva estatísticas para um conjunto de dados.
- ◆ **Titles**: especifica os títulos que aparecem no resultado.
- ◆ **Variables**: especifica peso, por grupo, e frequência de variáveis.

Exemplo 9.1.3

Consideremos os dados do exemplo 9.1.1

Analisando a Distribuição

Pode-se analisar as propriedades de distribuição dos dados com Distributions. Esta tarefa permite a você produzir estatística descritiva para as variáveis, testar a combinação de várias distribuições para os dados, e examinar dispositivos como por exemplo histogramas e distribuições de probabilidades. Nesta tarefa, o interesse está em examinar as taxas de natalidade e mortalidade infantis para cada região.

Solicitar uma Análise Distributiva (*Distributions Analysis*)

Para solicitar Distribuição, siga os seguintes passos:

1. Selecione **Statistics Descriptive** → **Distributions**.

2. Selecione *birth* e *death* como as variáveis de análise.
 3. Selecione *region* como a variável de classificação.
- A Figura 9.1.12 mostra o diálogo Distribuição com as especificações anteriores das variáveis.

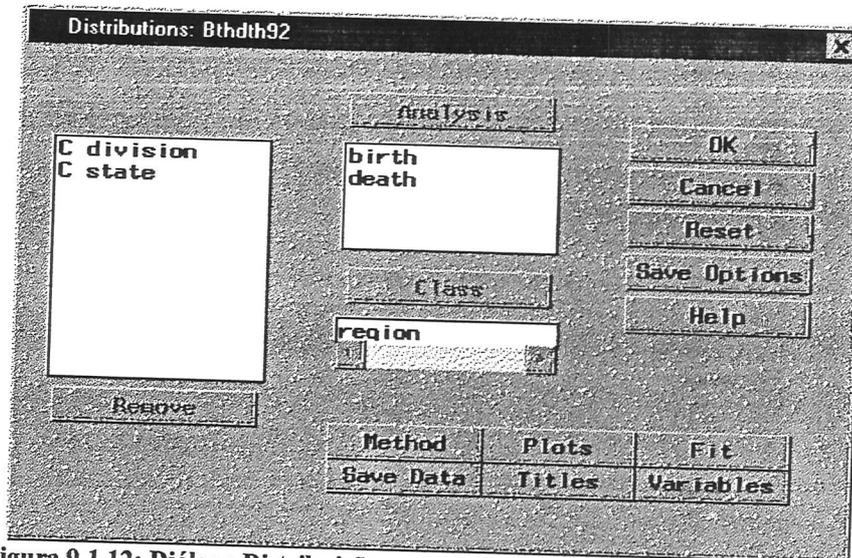


Figura 9.1.12: Diálogo Distribuição.

A análise padrão fornece momentos, quantis e medidas de variabilidade.

Solicitar Gráficos

Para solicitar box-and-whisker plots e histogramas, siga os seguintes passos:

1. Clique no botão **Plots**.
2. Selecione **Box-&-whisker plot**.
3. Selecione **Histogram**.
4. Clique **OK**.

A figura 9.1.13 mostra o diálogo Plots.

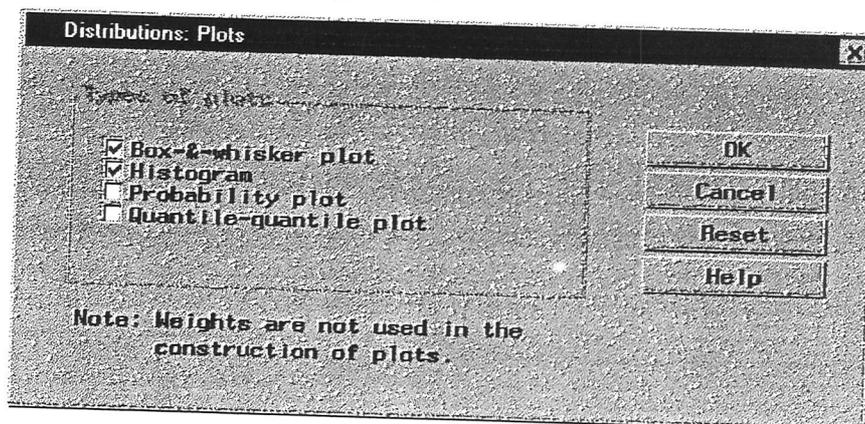


Figura 9.1.13: Distribuições: Diálogo Plots.

Solicitar Distribuição Ajustada

Para ajustar uma distribuição normal destes dados, siga os seguintes passos

1. Clique no botão **Fit** no diálogo principal.
2. Selecione **Normal**.

Por definição, os valores dos parâmetros são calculados a partir dos dados quando você ajusta a uma distribuição normal. Se você quiser entrar com parâmetros de cálculo específicos, clique na seta apontando para baixo (mostrada na figura 9.1.14) e selecione **Enter values**. Para as distribuições lognormal, exponencial, e Weibull, pode-se especificar que parâmetros sejam calculados pela estimação de máxima verossimilhança (MLE), ou você pode entrar com os parâmetros específicos para as variáveis.

3. Clique **OK**.

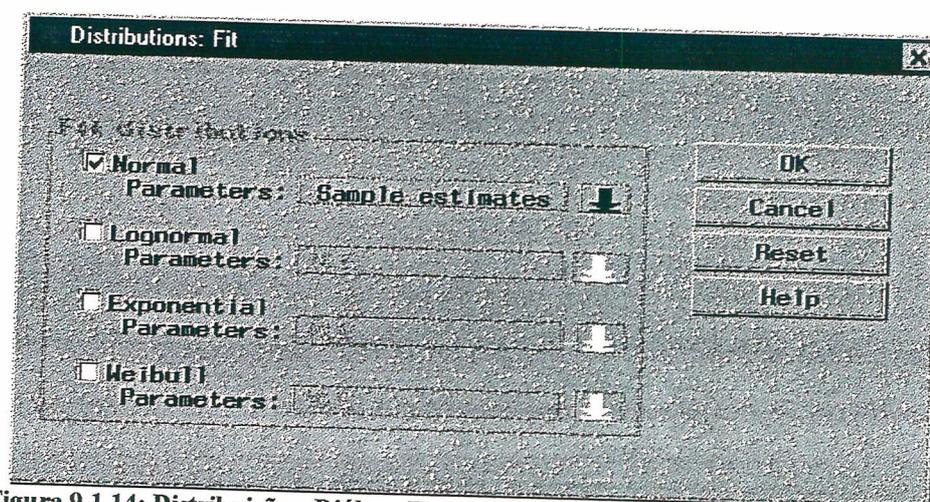


Figura 9.1.14: Distribuições: Diálogo Fit

Quando você tiver completado suas seleções, clique **OK** no diálogo principal para executar a análise. Os resultados são apresentados na árvore de projetos mostrada na Figura 9.1.15.

Análise dos Resultados (*Review the Results*)

Dê um duplo-clique em qualquer dos oito ícones de resultados para ver o correspondente resultado numa janela separada.

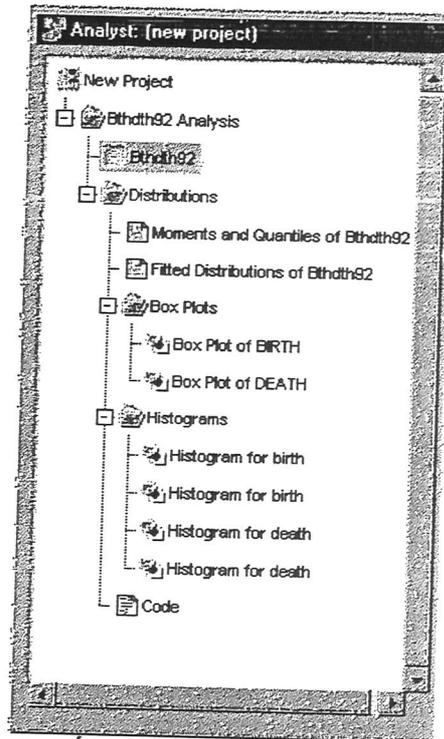


Figura 9.1.15: Distribuições: Árvore de Projeto.

Os resultados de Momentos e Quantis fornece informações resumidas para cada variável. A Figura 9.1.16, mostra o resultado chamado Fitted Distributions of Bthdth92, no qual resume de que forma a distribuição normal é combinada com cada variável, por região.

The UNIVARIATE Procedure
Fitted Distribution for birth
region = MW

Parameters for Normal Distribution

Parameter	Symbol	Estimate
Mean	Mu	14.825
Std Dev	Sigma	0.758138

Goodness-of-Fit Tests for Normal Distribution

Test	---Statistic---	-----p Value-----
Kolmogorov-Smirnov	D 0.14881321	Pr > D >0.150
Cramer-von Mises	W-Sq 0.04851780	Pr > W-Sq >0.250
Anderson-Darling	A-Sq 0.29379085	Pr > A-Sq >0.250

Figura 9.1.16: Distribuições: Resultados da Distribuição Combinada

Baseado no resultado mostrado na Figura 9.1.16, a hipótese de nulidade de que a variável *birth* é normalmente distribuída não pode ser rejeitada no nível de significância (p -

values para todos os testes são maiores que 0.15). O mesmo é verdadeiro para a variável *death*, exceto para a região Sul (region='S'). A hipótese é rejeitada no nível de significância para a taxa de mortalidade na região Sul.

Dois grupos de box plots e quatro grupos de histogramas também são gerados. Um único box-and-whisker plot é criado para cada uma das duas variáveis. O box-and-whisker plot para a variável *birth* é mostrado quando você dá um duplo clique em **Box Plot of BIRTH** na árvore de projeto.

Dois histogramas são criados para cada variável. Cada gráfico contém um histograma para dois níveis de classificação da variável *region*. O primeiro histograma contém as informações para as regiões Meio-Oeste e Nordeste (region='MW' e region='NE'), como mostrado na Figura 9.1.17. O segundo histograma (não mostrado) contém as informações das regiões Sul e Oeste. (region='S' e region='W').

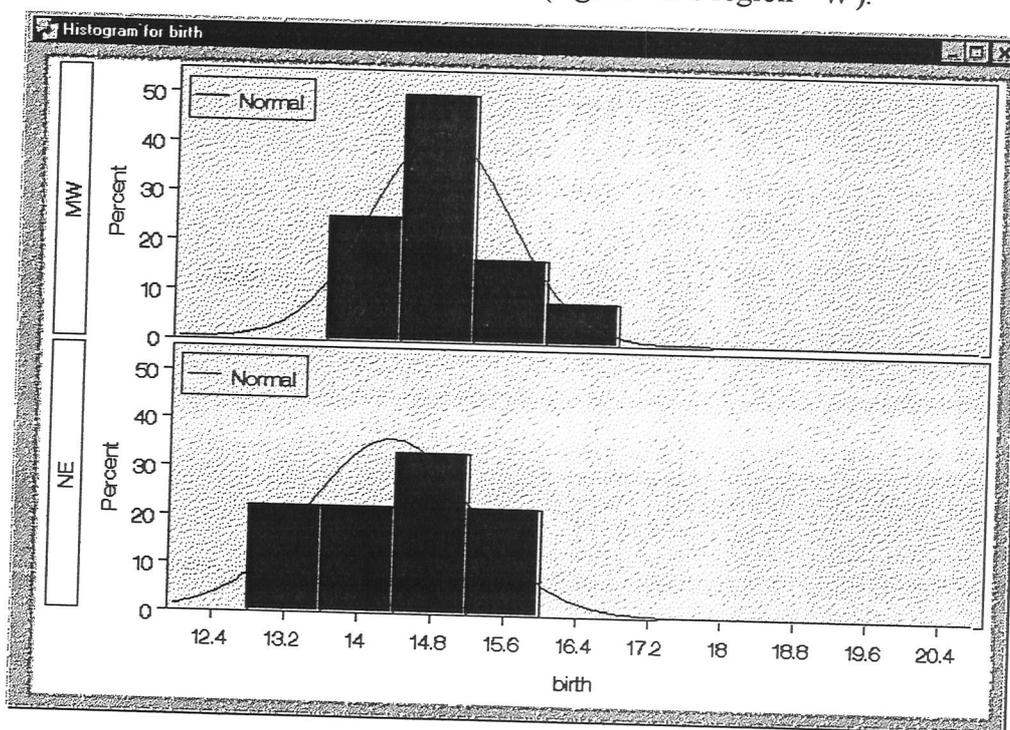


Figura 9.1.17: Distribuições: Histograma para *birth*

A curva normal sobre o histograma mostrado na figura 9.1.17 é o resultado do pedido de distribuição normal feito no diálogo Fit (Figura 9.1.14). Os detalhes estatísticos do fit estão localizados no output chamado Fitted Distributions of Bthdth92, o qual inclui também os detalhes do fit para a variável *death*.

Exercício 13.

Ainda com o banco de dados **pesovelo.xls** calcule se a variável altura se ajusta a distribuição normal.

9.1.5 Correlações

Selecione **Descriptive** do menu **Statistics**, então selecione **Correlations** para mostrar o diálogo **Correlations**. Você pode utilizar o diálogo **Correlations** para computar coeficientes de correlação para variáveis no conjunto de dados. A correlação entre duas variáveis é uma medida do grau de relacionamento linear entre as variáveis. Se um relacionamento linear exato existe, a correlação será 1 ou -1, dependendo se a inclinação do relacionamento é positiva ou negativa. Uma correlação de 0 indica que nenhuma variável tem qualquer habilidade preditiva para a outra (embora elas possam ter um relacionamento não linear). Se as variáveis são distribuídas normalmente e a correlação é zero, então elas são independentes. Você pode computar a correlação momento-produto de Pearson padronizada, como também as medidas de associação não paramétricas, correlações parciais, e o coeficiente alfa de Cronbach.

Lista de Candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibir na análise. Um "C" que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto a lista de candidatos, na qual todas as variáveis estão no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista por clique duplo nelas.

Botão Correlate

Selecione uma variável da lista de candidatos e clique no botão **Correlate** para colocar a variável na lista **Correlate**. A lista **Correlate** contém as variáveis que são analisadas. Correlações são computadas para cada par de variáveis na lista. Alternativamente, você pode dar um clique duplo na variável na lista de candidatos para colocá-la na lista **Correlate**.

Opções de Correlação

- Os seguintes diálogos estão disponíveis para personalizar sua análise.
- ◆ **Options**: especifica que tipos de correlação e estatísticas estão sendo produzidas.
 - ◆ **Plots**: produz gráficos de dispersão.
 - ◆ **Save Data**: salva correlações para um conjunto de dados.
 - ◆ **Titles**: especifica os títulos que aparecem nos resultados.
 - ◆ **Variables**: especifica peso, por grupo, frequência, e variáveis parciais.

Exemplo 9.1.4

O exemplo a seguir computa o coeficiente de correlação para quatro variáveis no banco de dados Fitness. Este banco de dados contém medidas obtidas de um grupo de homens de um curso de educação física na Universidade da Carolina do Norte. As variáveis são as seguintes:

age - idade, em anos

weight - peso, em quilogramas

oxygen - taxa de absorção de oxigênio, em milímetros por quilograma de peso corporal por minuto.

runtime - tempo para correr 1.5 milhas, em minutos.

rstpulse - batimento cardíaco em descanso.

Runpulse - batimento cardíaco enquanto corre.

maxpulse - máximo batimento cardíaco durante a corrida.

group - número do grupo.

Este exemplo engloba as correlações entre as variáveis *runtime*, *runpulse*, *maxpulse*, e *oxygen* e também produz o correspondente gráfico de dispersão com elipses de confiança.

Abrir (*Open*) o Banco de Dados Fitness (*Fitness Data Set*)

Para abrir o banco de dados Fitness, siga os seguintes passos:

1. Selecione **Tools** → **Sample Data ...**
2. Selecione **Fitness**.
3. Clique **OK** para criar o banco de dados exemplo no seu diretório Sasuser.
4. Selecione **File** → **Open By SAS Name ...**
5. Selecione Sasuser da lista de **Libraries**.
6. Selecione Fitness da lista de membros.
7. Clique **OK** para trazer o banco de dados Fitness para a tabela de dados.

Solicitar Correlações (*Correlations*)

Para computar correlações para variáveis no banco de dados Fitness, siga os seguintes passos:

1. Selecione **Statistics Descriptive** → **Correlations ...**
2. Selecione as variáveis *runtime*, *runpulse*, *maxpulse*, e *oxygen* para combinar.

A Figura 9.1.18 mostra o resultado do diálogo Correlações.

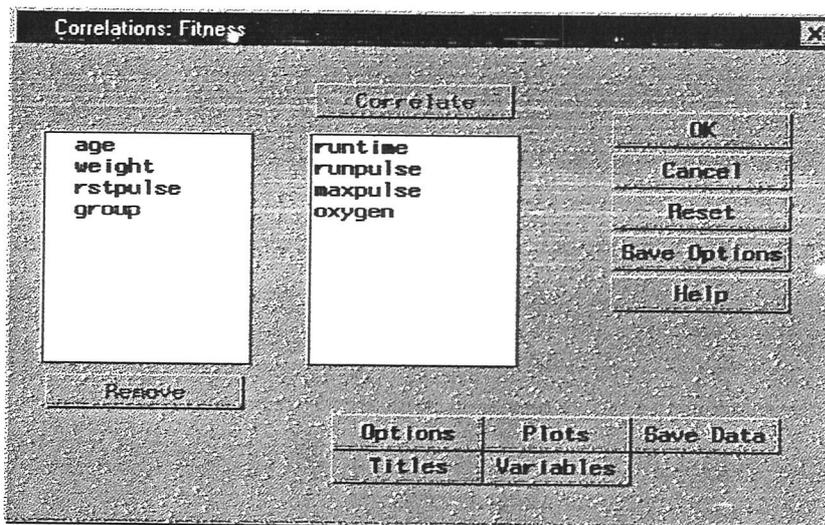


Figura 9.1.18: Diálogo Correlações

Se você clicar **OK** no diálogo principal Correlações, a saída padrão, que inclui as correlações de Pearson, é produzida. Senão, você pode pedir tipos específicos de correlações usando o diálogo Opções (Options).

Solicitar um Gráfico de Dispersão (*Scatter Plot*)

Para solicitar um gráfico de dispersão com uma elipse de confiança, siga os seguintes passos:

1. Clique no botão **Plots**.
2. Selecione **Scatter plots**.
3. Selecione **Add confidence ellipses**.

O nível de confiança usado para cálculo da elipse se confiança é 0.95. Para usar um nível diferente, digite o valor no campo **Probability value**, como mostrado na Figura 9.1.19.

4. Clique **OK**.

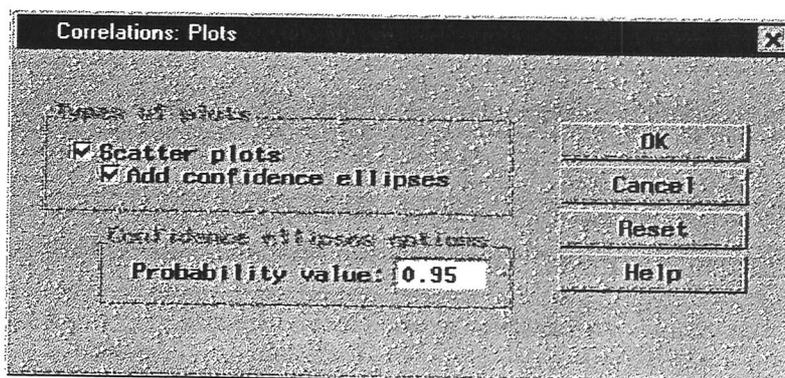


Figura 9.1.19: Correlações: Diálogo Plots

Clique **OK** no diálogo principal para proceder a análise.

Análise dos Resultados (*Review the Results*)

Os resultados são apresentados na árvore de projeto, como mostra a Figura 9.1.20.

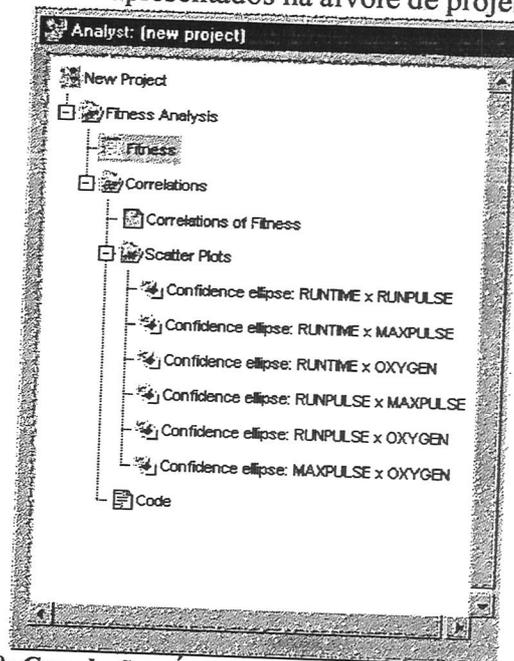


Figura 9.1.20: Correlações: Árvore de Projeto

Pode-se dar um duplo clique em qualquer dos nós de resultados na árvore de projeto para ver a informação numa janela separada.

A Figura 9.1.21 mostra estatísticas univariadas para cada uma das variáveis de análise. A tabela fornece o número de observações, a média, o desvio padrão, a soma e os valores mínimo e máximo para cada variável.

The screenshot shows the 'Correlations of Fitness' window with the following content:

The CORR Procedure
4 Variables: runtime runpulse maxpulse oxygen

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
runtime	31	10.58613	1.38741	328.17000	8.17000	14.03000
runpulse	31	169.64516	10.25199	5259	146.00000	186.00000
maxpulse	31	173.77419	9.16410	5387	155.00000	192.00000
oxygen	31	47.37581	5.32723	1469	37.38800	60.05500

Variable	Label
runtime	Min. to run 1.5 miles
runpulse	Heart rate while running
maxpulse	Maximum heart rate
oxygen	Oxygen consumption

Figura 9.1.21: Correlações: Estatísticas Univariadas

A Figura 9.1.22 mostra a tabela de correlações. O p -value, que é a probabilidade de significância da correlação, é mostrada abaixo de cada coeficiente de correlação. Por

exemplo, a correlação entre as variáveis *maxpulse* e *runtime* é 0.22610, com um *p*-value associado de 0.2213, e a correlação entre as variáveis *oxygen* e *runpulse* é -0.39797, com um *p*-value associado de 0.0266.

Correlations of Fitness

Pearson Correlation Coefficients, N = 31
Prob > |r| under H0: Rho=0

	runtime	runpulse	maxpulse	oxygen
runtime Min. to run 1.5 miles	1.00000	0.31365 0.0858	0.22610 0.2213	-0.86219 <.0001
runpulse Heart rate while running	0.31365 0.0858	1.00000	0.92975 <.0001	-0.39797 0.0266
maxpulse Maximum heart rate	0.22610 0.2213	0.92975 <.0001	1.00000	-0.23674 0.1997
oxygen Oxygen consumption	-0.86219 <.0001	-0.39797 0.0266	-0.23674 0.1997	1.00000

Figura 9.1.22: Correlações: Tabela de Correlações

Seis gráficos de dispersão, cada qual inclui uma elipse de confiança de 95%, são produzidos nesta análise. Cada gráfico mostra a relação entre um par de variáveis de análise. O gráfico de dispersão de *runtime* x *oxygen* está mostrado na Figura 9.1.23.

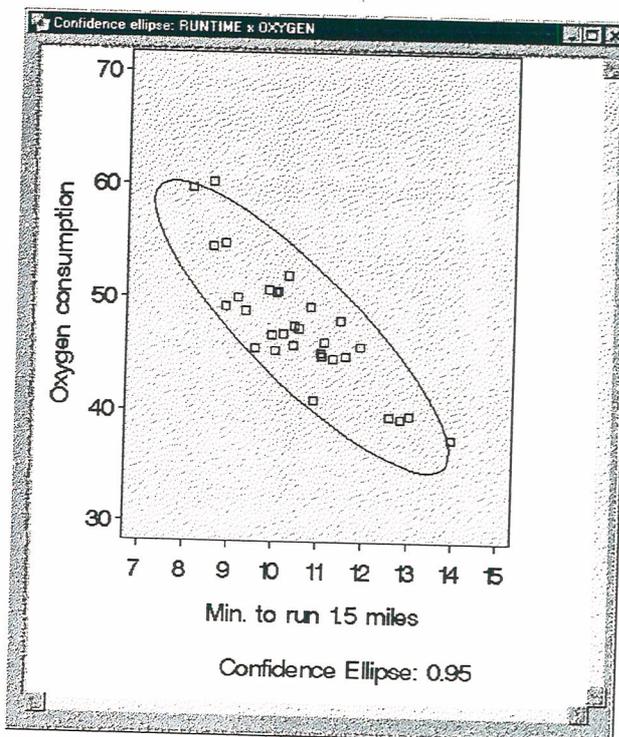


Figura 9.1.23: Correlações: Gráfico de Dispersão com Elipse de Confiança

Elipses de confiança são usadas como um indicador gráfico de correlação. Quando duas variáveis não são correlacionadas, a elipse de confiança tem forma circular. A elipse torna-se mais alongada quanto mais forte for a correlação entre as duas variáveis.

Exercício 14.

Agora calcule a correlação entre o **peso** e a **altura**.

9.2 Análise de Tabelas

9.2.1 Introdução

Freqüentemente necessita-se analisar informações dispostas em tabelas, comumente denominadas como tabelas de contingência ou de classificação cruzada. Pode-se analisar uma única tabela ou um conjunto de tabelas. Muitas vezes interessa avaliar a presença de associação na tabela, ou seja se existe algum tipo de relação entre as variáveis que caracterizam as linhas e as colunas da tabela. Várias estatísticas Qui-quadrado, tais como o Qui-quadrado de Pearson e da razão de verossimilhança, são utilizados para avaliar a associação.

Além de avaliar a presença de associação, pode também interessar o cômputo de uma medida de associação, ou de alguma estatística que forneça algum entendimento do grau ou magnitude da associação. A razão de chances("odds ratio") é uma medida de associação padrão e, tem sido usada com freqüência, em estudos em medicina e epidemiologia.

A tarefa **Análise de Tabelas** fornece o teste Qui-quadrado para tabelas $r \times c$, onde r é o número de linhas e c é o número de colunas da tabela, incluindo as estatísticas para o teste Qui-quadrado de Pearson e o teste da razão de verossimilhança e também fornece a extensão do teste de Mantel-Haenszel para conjuntos de tabelas. O teste exato de Fisher pode ser computado para tabelas 2×2 e $r \times c$. Adicionalmente a tarefa também permite a obtenção de medidas de associação, tais como os "odds ratio" e o risco relativo para tabelas 2×2 , bem como os coeficientes Gamma, Tau-b, D de Somer e os coeficientes de correlação de Pearson e Spearman. Permite também a obtenção de medidas de concordância, tais como os coeficientes kappa e kappa ponderado. O teste de MacNemar é executado para tabelas 2×2 .

9.2.2 Análise de Tabelas

Selecione **Table Analysis** no menu **Statistics** para exibir o diálogo de análise de tabelas. Neste diálogo, você pode criar e analisar tabelas de contingência com duas até n variáveis de classificação. Você deve especificar no mínimo uma variável para a classificação referente as linhas e no mínimo uma variável para a coluna. Você pode também especificar estratos de variáveis.

O menu para a seleção da Análise de Tabelas aparece na figura 9.2.1

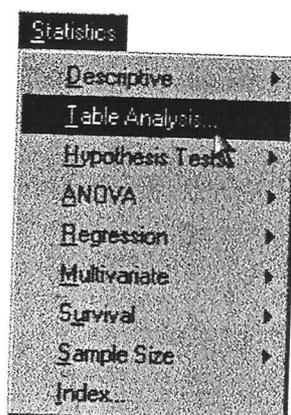


Figura 9.2.1 Menu para a seleção da Análise de Tabelas

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las na análise. Um “C” que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Remove

O botão **Remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Botão Row

Selecione variáveis da lista de candidatos e clique no botão **Row** para especifica-las como variáveis linhas.

Botão Column

Selecione variáveis da lista de candidatos e clique no botão **Column** para especifica-las como variáveis colunas.

Botão Strata

Selecione variáveis da lista de candidatos e clique no botão **Strata** para especifica-las como variáveis estratificadoras.

Botão Cell Counts

Selecione uma variável da lista de candidatos e clique no botão **Cell Counts** para especificar se ela é uma variável de contagem. A variável de contagem permite que cada observação no conjunto dos dados represente mais ou menos um sujeito. Por exemplo, se o valor da variável de contagem é quatro para uma observação em particular, então esta observação representa quatro sujeitos. Os valores das variáveis de contagem não tem que ser inteiros. Se os valores da variável de contagem são zero ou ausente (missing”), a observação correspondente é ignorada. Se o valor é negativo, as frequências são impressas, mas outras estatísticas são suprimidas.

Opções de Análises de Tabelas

Os seguintes diálogos estão disponíveis para personalizar suas análises.

- **Input:** Especifica a ordem dos valores da variável.
- **Select Tables:** Escolhe qual tabela incluir na análise.
- **Statistics:** Especifica as estatísticas a serem produzidas.
- **Tables:** Especifica a informação a incluir nas tabelas.
- **Titles:** Especifica os títulos que aparecem nos resultados.
- **Variables:** Especifica POR grupo de variáveis

9.2.3 Associação em Tabelas 2 x 2

A tabela de contingência mais básica é a tabela 2 x 2. Nesta tabela, usualmente, as colunas representam algum tipo de resultado, muitas vezes sim e não, e as linhas representam os níveis de um fator(variável) que pode ter influência no resultado.

Exemplo 9.2.1

Suponha, por exemplo, que pesquisadores estariam investigando as propriedades de um novo bandaid "ouchless" para crianças. Posições interessantes em que as crianças testaram o bandaid mostraram menos reclamações na retirada do que as crianças que utilizaram um bandaid normal. Você pode investigar esta questão criando a tabela two-way de tipo de bandaid e status de reclamação e então avaliar a associação entre as linhas e colunas desta tabela.

Abrindo o Conjunto de Dados Bandaid

Estes dados são fornecidos como o conjunto de dados Bandaid na Analyst Sample Library. Para abrir o conjunto de dados Bandaid, siga estes passos:

1. Selecione **Tolls** → **Sample Data ...**
2. Selecione Bandaid.
3. Clique **OK** para criar o conjunto de dados amostrais no seu diretório Sansuser.
4. Selecione **File** → **Open By SAS Name ...**
5. Selecione Sasuser da lista de **Libraries**.
6. Selecione Bandaid da lista de membros.
7. Clique **OK** para trazer o conjunto de dados Bandaid para a tabela de dados.

A figura 9.2.2 mostra a tabela de dados contendo estes dados. Note que os dados estão em forma de frequência, com a variável contagem contendo as frequências dos perfis contidos em cada linha da tabela. O tipo de variável é o tipo de bandaid testado e a variável resultado é o status de reclamação.

Bandaid (Browse)			
	type	outcome	count
1	regular	complain	14
2	regular	no	16
3	test	complain	10
4	test	no	30

Figura 9.2.2: Conjunto de Dados Bandaid na Tabela de Dados

Especificando a Tabela

Para construir a tabela two-way apropriada e solicitar testes de associação, siga estes passos:

1. Selecione **Statistics** → **Table Analysis ...**
2. Selecione o tipo da lista de candidatos como a variável **Row**.
3. Selecione resposta da lista de candidatos como a variável **Column**.
4. Selecione contagem da lista de candidatos como a variável **Cell Counts**.

A figura 9.2.3 mostra o diálogo resultante.

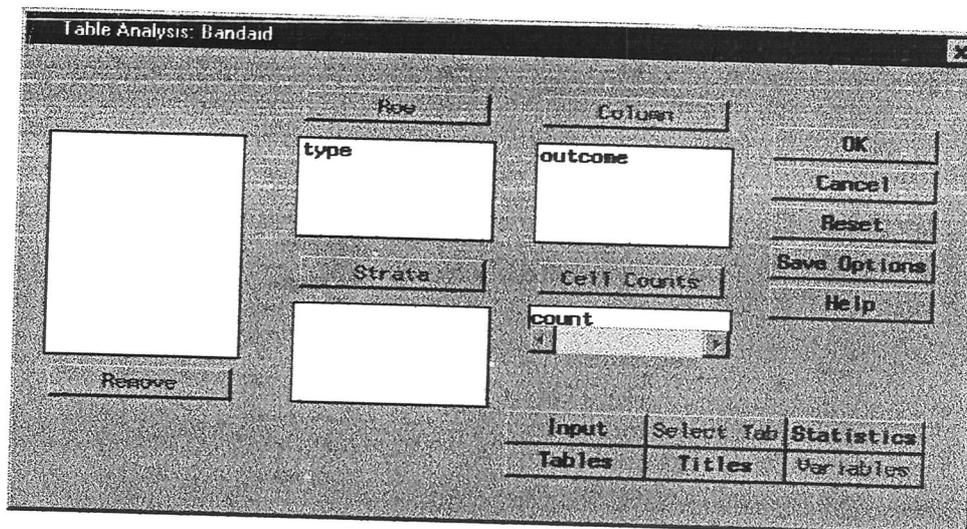


Figura 9.2.3: Análise de Tabela para o Estudo Bandaid

Solicitando Testes e Medidas de Associação

Selecionando as linhas e as colunas da tabela, você solicita a construção de uma tabela 2 x 2. Para solicitar testes qui-quadrado de associação razões de chances, a qual é uma medida de associação, siga estes passos:

1. Clique no botão **Statistics**.
2. Selecione **Chi-square statistics**.
3. Selecione **Measures of association**.
4. Clique **OK**.

A figura 9.2.4 mostra o diálogo **Statistics**.

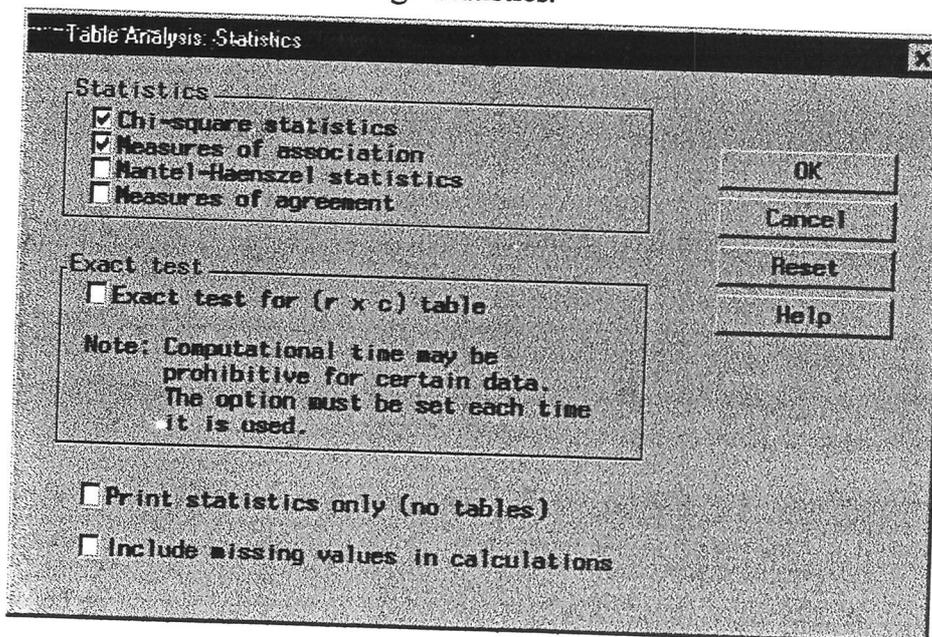


Figura 9.2.4: Diálogo Statistics

Finalmente, em ordem de personalizar a forma da tabela mostrada, siga estes passos:

1. Clique no botão **Tables**.
2. Selecione **Observed** abaixo **Frequencies**.
3. Selecione **Row** abaixo **Percentages**.
4. Clique **OK**.

A figura 9.2.5 mostra o diálogo Tables resultante.

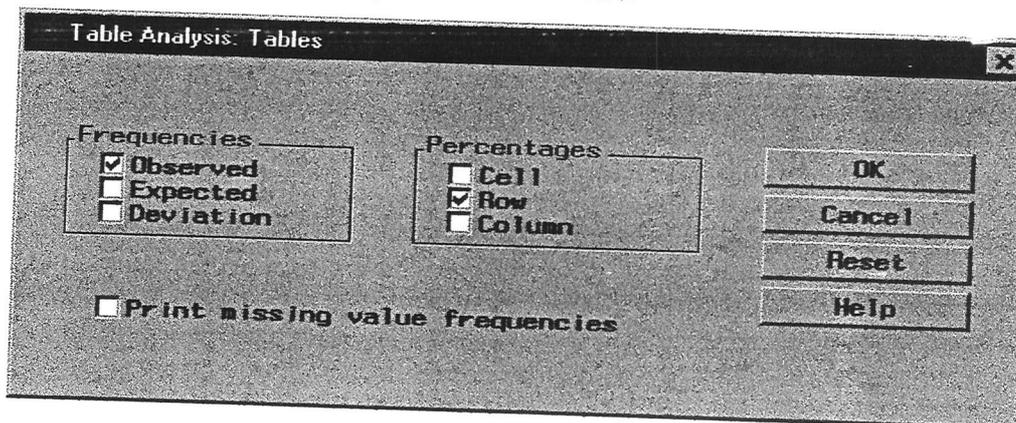


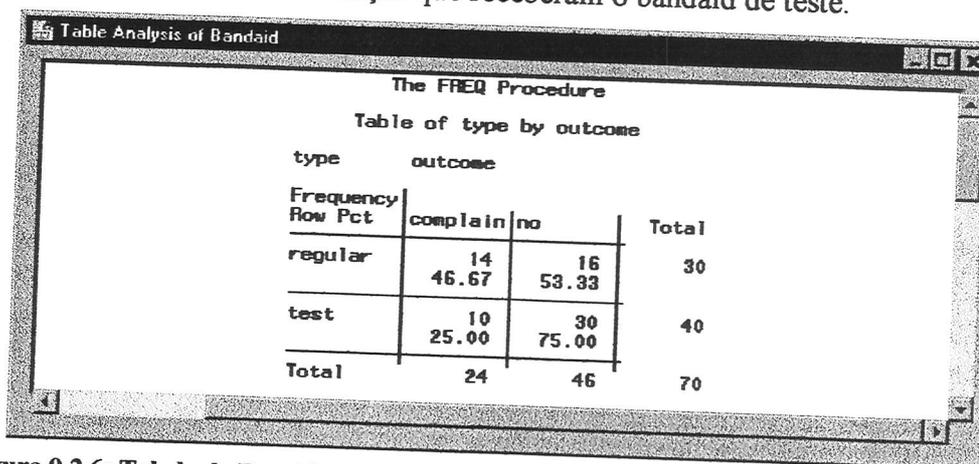
Figura 9.2.5: Diálogo Tables

Esta requer que somente as frequências de linhas e as porcentagens das linhas sejam listadas na célula da tabela impressa.

Clique **OK** no diálogo Table Analysis para executar a análise.

Examinando os Resultados

A tabela de frequência é mostrada na figura 9.2.6. Note que 46% das crianças que receberam bandaid's normais tinham reclamações sobre irritação quando seus bandaid's foram retirados, em relação a 25% das crianças que receberam o bandaid de teste.



type	outcome		Total
	complain	no	
regular	14 46.67	16 53.33	30
test	10 25.00	30 75.00	40
Total	24	46	70

Figura 9.2.6: Tabela de Frequência para Dados de Bandaid

A figura 9.2.7 contém a tabela de estatísticas qui-quadrados computadas para esta tabela. A estatística qui-quadrado de Pearson, designada "Qui-Quadrado", tem um valor de 3.57 e um

p -value associado de 0.0588 com 1 grau de liberdade. Se você estava testando hipóteses rígidas, você não deveria rejeitar a hipótese de não associação para o nível de significância $\alpha = 0.05$. Embora neste caso, os pesquisadores encontraram evidências suficientes neste estudo piloto, continuaram investigando o novo produto.

Statistics for Table of type by outcome

Statistic	DF	Value	Prob
Chi-Square	1	3.5719	0.0588
Likelihood Ratio Chi-Square	1	3.5655	0.0590
Continuity Adj. Chi-Square	1	2.6749	0.1019
Mantel-Haenszel Chi-Square	1	3.5208	0.0606
Phi Coefficient		0.2259	
Contingency Coefficient		0.2203	
Cramer's V		0.2259	

Fisher's Exact Test

Cell (1,1) Frequency (F)	14
Left-sided Pr <= F	0.9840
Right-sided Pr >= F	0.0512
Table Probability (P)	0.0351
Two-sided Pr <= P	0.0769

Figura 9.2.7: Estatísticas Qui-Quadrado para Dados de Bandaid

Diversas outras estatísticas qui-quadrado também aparecem neste resultado, assim como o qui-quadrado da razão da verossimilhança e o qui-quadrado de Mantel-Haenszel. Estas estatísticas são assintoticamente equivalentes.

Estimates of the Relative Risk (Row1/Row2)

Type of Study	Value	95% Confidence Limits	
Case-Control (Odds Ratio)	2.6250	0.9530	7.2307
Cohort (Col1 Risk)	1.8667	0.9656	3.6085
Cohort (Col2 Risk)	0.7111	0.4865	1.0394

Sample Size = 70

Figura 9.2.8: Razão de Chances para os Dados de Bandaid

A figura 9.2.8 contém a tabela de estimativas de risco relativo incluindo a razão de chances, a qual é rotulada "Case-Control." A razão de chances é a razão das chances de haver uma resposta para um grupo versus outro. Quando a razão de chances tem o valor 1, você tem chance iguais de obter a resposta. Quando a razão das chances é maior que 1, um grupo tem mais chance de uma resposta que outro.

A razão de chances tem um valor de 2.62, o qual significa que as chances de uma reclamação são 2.62 vezes maiores para aquelas crianças que utilizaram o bandaid regular que para aquelas que utilizaram o bandaid do teste.

Exercício 15.

Faça uma análise de tabelas a partir dos dados do arquivo **broca.xls**.

9.3 Teste de Hipóteses

9.3.1 Introdução

Testes de hipóteses são freqüentemente executados para uma e duas amostras. Para uma amostra é muitas vezes de interesse verificar se uma característica da população, tal como a média é equivalente a um certo valor. Para duas amostras, pode ser do interesse verificar se as médias verdadeiras são diferentes. Para amostras pareadas, pode interessar se a verdadeira diferença média é nula.

O Analyst permite executar testes de hipóteses para médias, proporções e variâncias para uma e duas amostras. O Menu para teste de hipóteses aparece na figura 9.3.1.

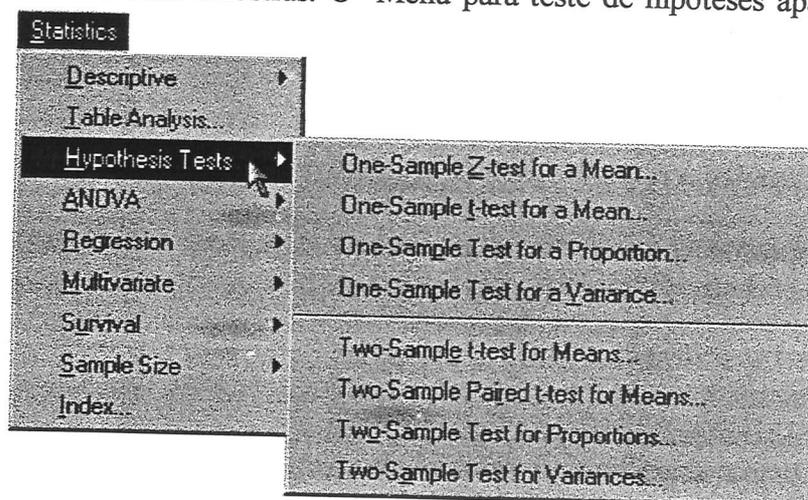


Figura 9.3.1 Menu para teste de hipóteses

9.3.2 Teste t para Média de uma Amostra

Selecione **Hypotesys Tests** do menu **Statistics** e então selecione **One-Sample t-test for a Mean** para exibir o diálogo do teste t para a média de uma amostra. Você pode testar se a média de uma população é igual ao valor especificado pela hipótese de nulidade. Este teste é apropriado quando você não sabe o desvio padrão da população de que seus dados pertencem. Se o desvio padrão da população é conhecido, então você pode usar o teste Z para a média de uma amostra. Na tarefa teste t para uma amostra, você deve especificar uma variável para ser analisada bem como uma média. A média da variável no conjunto de dados é comparada com a especificada na Hipótese.

Resultado do Teste-t para uma amostra

Os resultados padrões incluem um resumo estatístico para cada variável selecionada bem como a estatística t e o p-value. A estatística t mede o número de desvios padrões da média amostral distantes da média que você especificou na hipótese de nulidade. Calcula-se

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$$

onde s que é o desvio padrão amostral das variáveis selecionadas e n é o tamanho da amostra. O p-value é a probabilidade, se a hipótese de nulidade é verdadeira, de se obter um resultado que é no mínimo tão extremo quanto o que é obtido pelos dados.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las na análise. Um “C” que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Variable

A caixa **Variable** contém a variável que você deseja analisar. Selecione a variável da lista de candidatos e clique no botão **Variable**, para usá-la como uma variável. Alternativamente, um duplo clique sobre a variável na lista de candidatos a coloca na caixa.

Botão Remove

O botão **Remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Hipóteses de Nulidade e Alternativa

O teste de hipóteses executado na tarefa fornece um método para decidir qual das duas teorias sobre a verdadeira média das variáveis selecionadas é suportada pelos dados. A primeira teoria, conhecida como hipótese de nulidade, representa o status quo. Esta hipótese é preservada a menos que os dados dêem forte evidência contrária. Você deve especificar um valor da média para representar o status quo. A hipótese alternativa representa uma mudança de status quo. Você deve especificar uma das três escolhas disponíveis. A primeira escolha é que a média das variáveis não é igual a média especificada. Isto é, uma alternativa bilateral. As alternativas unilaterais são que a média das variáveis são maiores que ou menores que a média especificada. Note que a escolha da hipótese alternativa deveria ser baseada na expectativa do experimento, não sobre os dados. Quando o p-value é menor que uma probabilidade mínima (tipicamente 0,05 ou 0,01), nós “rejeitamos a hipótese de nulidade”. Isto é, nós concluímos que existe uma forte evidência a favor da hipótese alternativa especificada.

Hipótese de Nulidade

Especifica um valor para a hipótese de nulidade.

Hipótese Alternativa

Selecione uma hipótese alternativa. As três escolhas são **diferente, maior que, e menor que** o valor da hipótese de nulidade.

Opções do Teste-t para uma amostra

Os seguintes diálogos estão disponíveis para personalizar a sua análise.

- Tests: Produz intervalos de confiança ou uma análise do poder do teste.
- Plots: Especifica três tipos de Gráficos.

- Titles: Especifica os títulos que aparecem nos resultados.
- Variables: Especifica um grupo de variável.

Exemplo 9.3.1

Sejam os dados do exemplo 9.1.1. Suponha que você deseja determinar se a taxa de mortalidade infantil média nos Estados Unidos é igual a um valor específico. Note que o teste t one-sample é apropriado nesta situação porque o desvio padrão da população da qual os dados provém é desconhecido. Quando você conhece o desvio padrão da população, utilize o Teste Z One-Sample.

Abrindo o Conjunto de Dados Bthdth92

Proceda como no exemplo 9.1.1

Solicitando um Teste t One-Sample

Para testar se a taxa de mortalidade infantil média é igual à 8, siga estes passos:

1. Selecione **Statistics** → **Hypothesis Tests** → **One-Sample t-Test for a Mean ...**
2. Selecione morte como a variável a ser analisada.
3. Entre com **8** na caixa rotulada **Null: Mean =** e precione **Enter**.

Sua hipótese alternativa pode ser que a média seja menor que, maior que, ou diferente de um valor específico. Neste exemplo, a hipótese alternativa é que a média da variável morte seja diferente de 8.

Na figura 9.3.2, o diálogo teste t one-sample define as hipóteses nula e alternativa e especifica morte como a variável a ser testada.

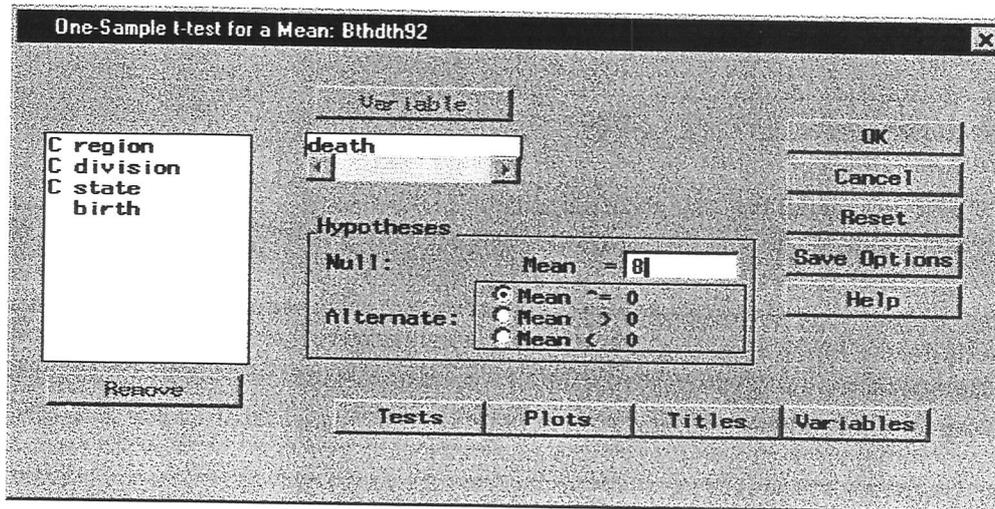


Figura 9.3.2: Diálogo Test-t One-Sample

A definição da tarefa *test-t* one-sample inclui estatísticas amostrais para a variável morte e os resultados do teste de hipóteses.

Computando um Intervalo de Confiança para a Média

Para produzir um intervalo de confiança para a média em adição ao teste de hipótese, siga estes passos:

1. Clique no botão **Tests** no diálogo principal.
2. Selecione **Interval** para solicitar um intervalo de confiança bilateral para a média

Você pode escolher um intervalo de confiança unilateral ou bilateral para a média. As seleções **Lower bound** e **Upper bound** especificam limites de confiança unilaterais.

Por definição, o nível de confiança é 95%. Você pode clicar na seta para baixo para selecionar outro nível de confiança, ou você pode colocar um nível de confiança na caixa.

3. Clique **OK** para retornar ao diálogo principal.

A figura 9.3.3 mostra a seleção de um intervalo de confiança bilateral de 95% para a média. Note que você também pode solicitar uma análise de poder retrospectiva do teste na placa **Power Analysis**.

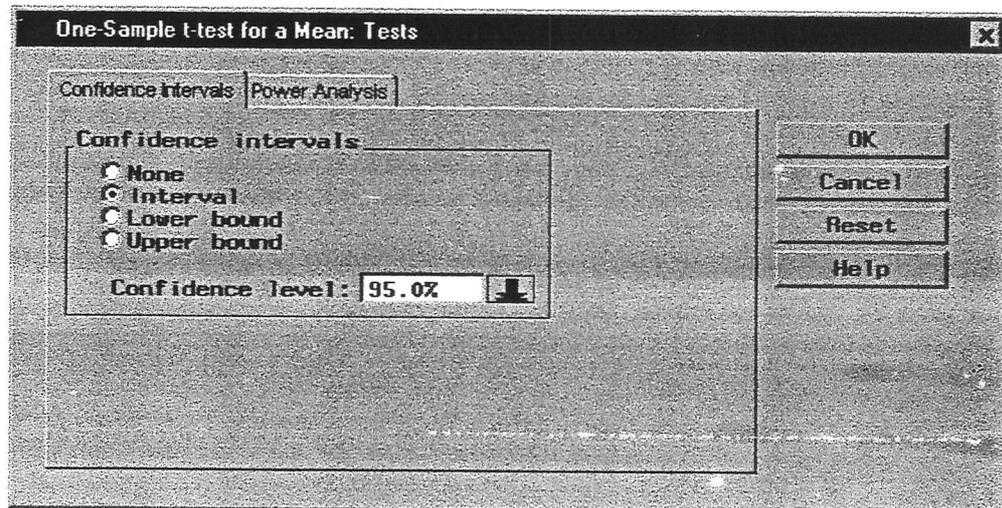


Figura 9.3.3: Teste t One-Sample: Diálogo Tests

Solicitando um Gráfico da Distribuição t

Para solicitar um gráfico da distribuição t em adição para o teste de hipótese, siga estes passos:

1. Clique no botão **Plots** no diálogo principal.
2. Selecione **t distribution plot**.
3. Clique **OK** para retornar ao diálogo principal.

A figura 9.3.4 mostra o diálogo Plots com **t distribution plot** selecionado.

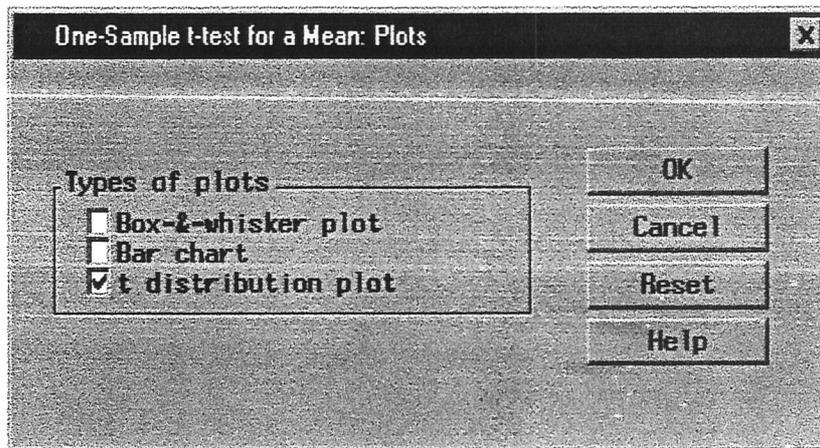


Figura 9.3.4: Teste t One-Sample: Diálogo Plots

Clique **OK** no diálogo principal para executar a análise.

Examinando os Resultados

Os resultados do teste de hipótese são mostrados na figura 9.3.5. O resultado inclui a tabela “Sample Statistics” para a variável morte, os resultados dos testes de hipóteses, e o intervalo de confiança de 95% para a média.

A média da variável morte é 8.61, a qual é maior que o valor de teste especificado 8.

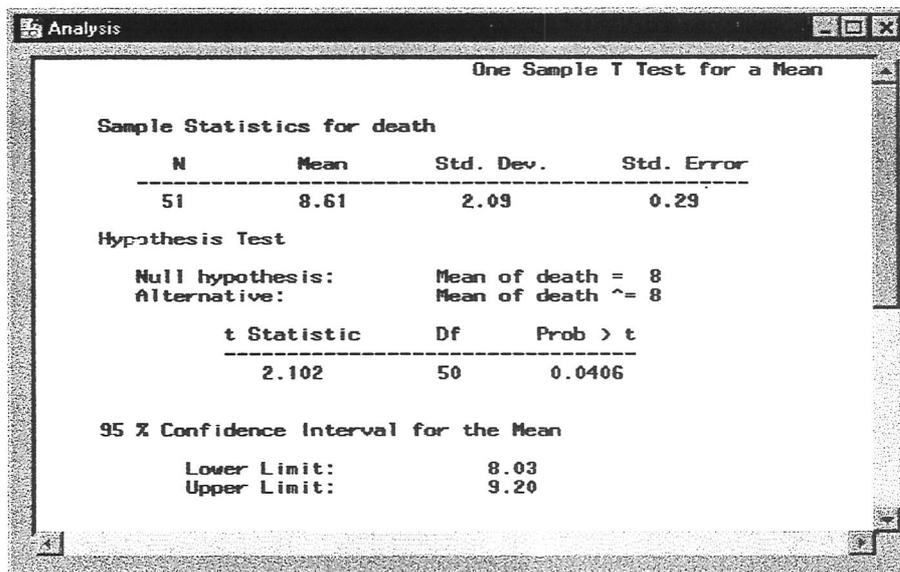


Figura 9.3.5: Teste t One-Sample: Resultados

A estatística t de 2.102 e o p -value associado (0.0406) fornecem evidências ao nível $\alpha = 0.05$ de que o índice de mortalidade infantil médio não é igual a 8. O intervalo de confiança indica que você pode estar 95% confiante de que a verdadeira média está dentro do intervalo [8.03, 9.20]. O gráfico de distribuição t solicitado é mostrado na

figura 9.3.6. O gráfico descreve a estatística t calculada sobreposta em uma função densidade de distribuição t com 50 graus de liberdade.

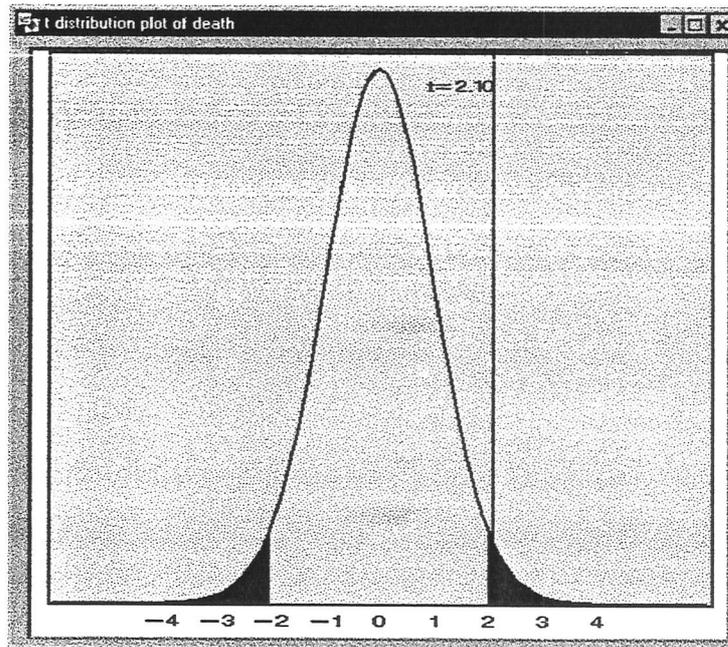


Figura 9.3.6: Teste t One-Sample: Gráfico da Distribuição t

Porque esta análise requer um teste bilateral, duas regiões críticas são sombreadas, uma em cada uma das caudas direita e esquerda. O nível alfa para o teste é 0.05; deste modo, cada região representa 2.5% da área abaixo da curva. Em um teste unilateral ao nível $\alpha = 0.05$, a região crítica aparece somente em uma cauda, e isto representa 5% da área abaixo da curva.

Aqui, a estatística t diminui na região sombreada. Desta forma, a hipótese nula é rejeitada.

Exercício 16.

Faça um teste- t para a hipótese utilizando os dados do arquivo **pesovelo.xls**.

9.3.3 Teste t para Duas Amostras Pareadas

Selecione **Hipotesys Test** do menu, então selecione **Two-Sample Paired t -test for Means** para exibir o diálogo para o teste t para duas amostras pareadas. Você pode testar se as médias de duas populações diferem de uma quantidade especificada. Dados pareados aparecem quando duas amostras dependentes são observadas. Isto pode ocorrer, por exemplo, quando duas medidas são tomadas (ou tratamentos são aplicados) em cada sujeito num estudo. Na tarefa Teste t para amostras pareadas, você deve especificar as duas amostras selecionando uma variável para a caixa **Group 1** e uma para o **Group 2**. A hipótese de teste será realizada sobre a diferença das médias das duas amostras, e você deve especificar um valor para esta diferença. Por definição, o valor é zero, que resulta num teste em que as médias das duas populações são iguais. Você pode também escolher uma das três

hipóteses alternativas. Os resultados comparam a diferença especificada com as diferenças entre as médias amostrais.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibí-las na análise. Um “C” que precede a variável indica que ela é uma variável não numérica (caracter). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Group1

Selecione a variável contendo os dados e clique no botão **Group 1** para posicionar a variável na caixa.

Botão Group 2

Selecione a variável contendo os dados e clique no botão **Group 2** para posicionar a variável na caixa.

Botão Remove

O botão **Remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Hipóteses de Nulidade e Alternativa

O teste de hipótese executado na tarefa fornece um método para decidir qual das duas teorias sobre a verdadeira diferença entre as médias das variáveis selecionadas é suportada pelos dados. A primeira teoria, conhecida como hipótese de nulidade, representa o status quo. Esta hipótese é preservada a menos que os dados dêem forte evidência contrária. Você deve especificar um valor para a diferença entre as médias para representar o status quo. A hipótese alternativa representa uma mudança de status quo. Você deve especificar uma das três escolhas disponíveis. A primeira escolha é que a diferença não é igual à diferença especificada. Isto é, uma alternativa bilateral. As alternativas unilaterais são que a média de diferenças entre as variáveis é maior que ou menor que a diferença média especificada. Note que a escolha da hipótese alternativa deveria ser baseada na expectativa do experimento, não sobre os dados. Quando o p-value é menor que uma probabilidade mínima (tipicamente 0,05 ou 0,01), nós “rejeitamos a hipótese de nulidade”. Isto é, nós concluímos que há forte evidência favorável a hipótese alternativa.

Hipótese de Nulidade

Especifica um valor para a hipótese de nulidade da diferença entre as médias. Especialmente teste-se se as médias são iguais, isto é se a diferença é igual a zero.

Hipótese Alternativa

Selecione uma hipótese alternativa. As três escolhas são **diferente, maior que, e menor que** o valor da hipótese de nulidade.

Opções do Teste-t para Amostras Pareadas

Os seguintes diálogos estão disponíveis para personalizar a sua análise.

- **Tests:** Produz intervalos de confiança ou análise do poder do teste.
- **Plots:** Especifica quatro tipos de Gráficos.
- **Titles:** Especifica os títulos que aparecem nos resultados.
- **Variables:** Especifica um grupo de variável.

Exemplo 9.3.2

Por exemplo, os dados de natalidade e morte analisados no exemplo 9.3.1 da seção anterior são considerados relacionados ou pareados porque, em cada observação, as variáveis natalidade e morte correspondem ao mesmo estado.

Suponha que você deseja determinar se as médias para a taxa de natalidade e para a taxa de mortalidade infantil são iguais. O Analyst fornece o teste t um teste para Two-Sample pareadas para a tarefa Means, o qual testa a igualdade de médias de duas amostras pareadas ou relacionadas. As duas amostras neste exemplo são a taxa de natalidade (natalidade) e a taxa de mortalidade infantil (morte) para cada estado.

Abrindo o conjunto de Dados Bthdth92

Proceda como no exemplo 9.1.1

Solicitando um Teste t para amostras Pareadas

Para executar esta análise, siga estes passos:

1. Selecione **Statistics** → **Hypothesis Tests** → **Two-Sample Paired t-test for Means ...**
2. Selecione a variável natalidade como a variável do Grupo 1.
3. Selecione a variável morte como a variável do Grupo 2.

O teste de interesse é se a diferença das médias é zero. Este é, por definição, o valor no Analyst, embora você possa especificar outros valores também.

Você pode escolher uma de três hipóteses alternativas. A definição é que a diferença entre médias não seja igual a diferença especificada, a qual é a alternativa bilateral. As alternativas unilaterais são que a diferença é maior que, ou menor que, a diferença especificada na hipótese de nulidade.

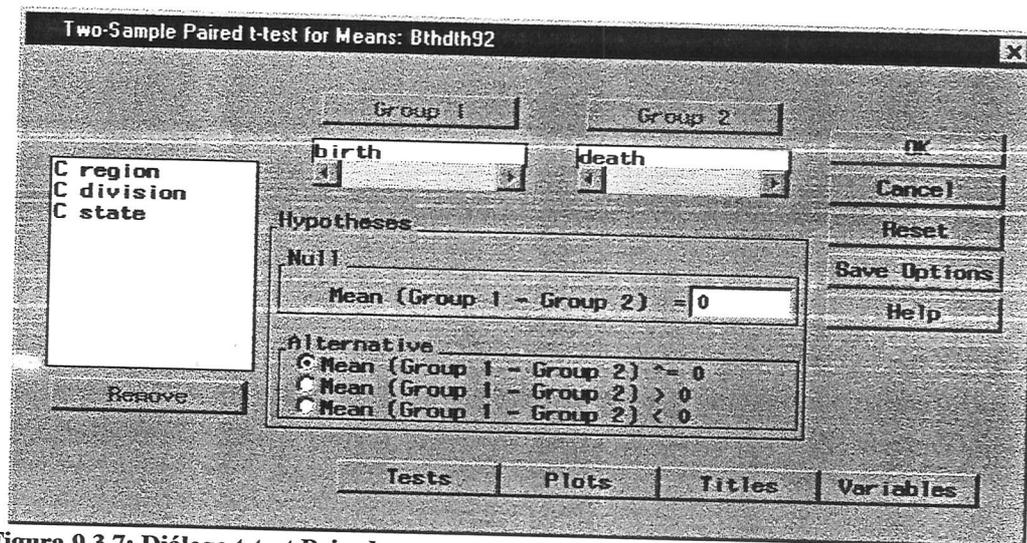


Figura 9.3.7: Diálogo t-test Paired

Na figura 9.3.7, a hipótese nula especifica que as médias das variáveis natalidade e morte são iguais (ou, equivalente, que a diferença entre as médias é zero). A hipótese alternativa é que as duas médias não são iguais.

Solicitando Gráficos

Para especificar um gráfico box-and-whisker e um gráfico das médias em adição ao teste de hipótese, siga estes passos:

1. Clique no botão **Plots** no diálogo principal.
2. Selecione **Box-&-whisker plot**.
3. Selecione **Means plot**.
4. Clique **OK**.

A figura 9.3.8 mostra o diálogo Plots com **Box-&-whisker plot** e **Means plot** selecionados.

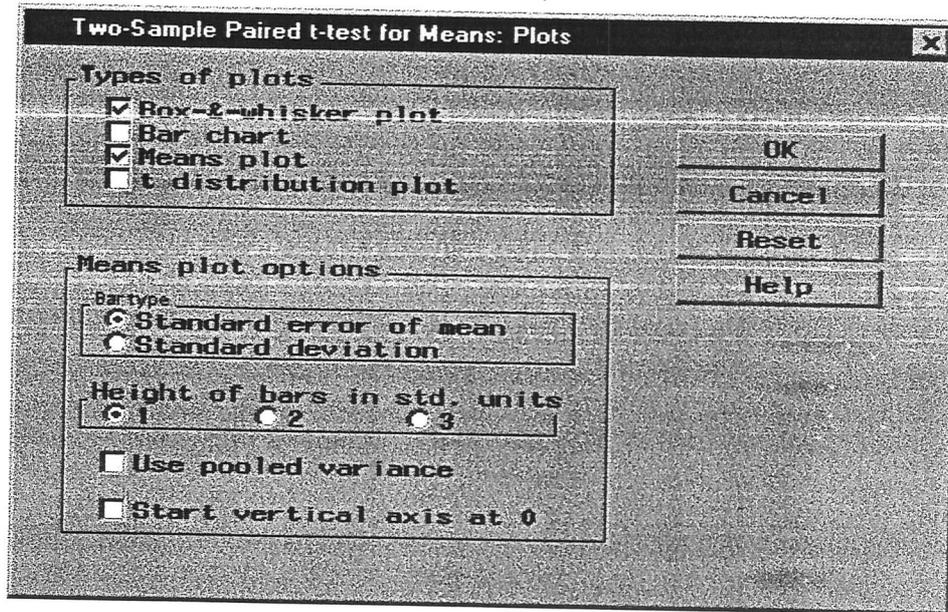


Figura 9.3.8: Teste t para amostras pareadas: Diálogo Plots

Clique **OK** no diálogo principal para executar a análise.

Examinando os Resultados

Os resultados das análises, mostrados na figura 9.3.9, contêm a média, desvio padrão e erro padrão da média para ambas as variáveis. A tabela "Hypotesis Test" fornece a estatística t observada, os graus de liberdade, e o p -value associado ao teste

Two Sample Paired t-test for the Means of birth and death				
Sample Statistics				
Group	N	Mean	Std. Dev.	Std. Error
birth	51	15.48431	1.7202	0.2409
death	51	8.613725	2.0851	0.292

Hypothesis Test		
Null hypothesis:	Mean of (birth - death) = 0	
Alternative:	Mean of (birth - death) \neq 0	
t Statistic	Df	Prob > t
19.926	50	<.0001

Figura 9.3.9: Teste t para amostras pareadas: Resultados

Na figura 9.3.9, a tabela "Sample Statistics" mostra que a média da variável natalidade é maior que da variável morte. Na tabela "Hypothesis Test", a estatística t (19.926) e o p -value associado (< 0.0001) indicam que a diferença entre as duas médias é estatisticamente muito significativa.

A figura 9.3.10 mostra o box-plot lado a lado de natalidade e morte. Observações que caem além das linhas são individualmente identificadas com um símbolo quadrado.

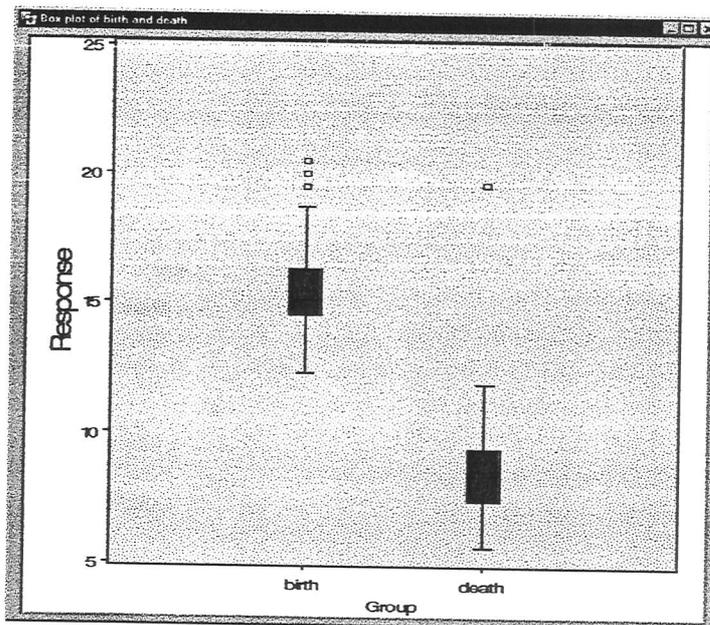


Figura 9.3.10: Teste t para amostras pareadas: Gráfico Box-and-Whisker

Os gráficos de média e erro padrão mostrados na figura 9.3.11 fornecem uma outra visão das duas variáveis. O gráfico de médias descreve um intervalo centralizado na média amostral para cada variável. O intervalo da linha vertical estende-se em dois desvios padrões em cada um dos dois lado da média.

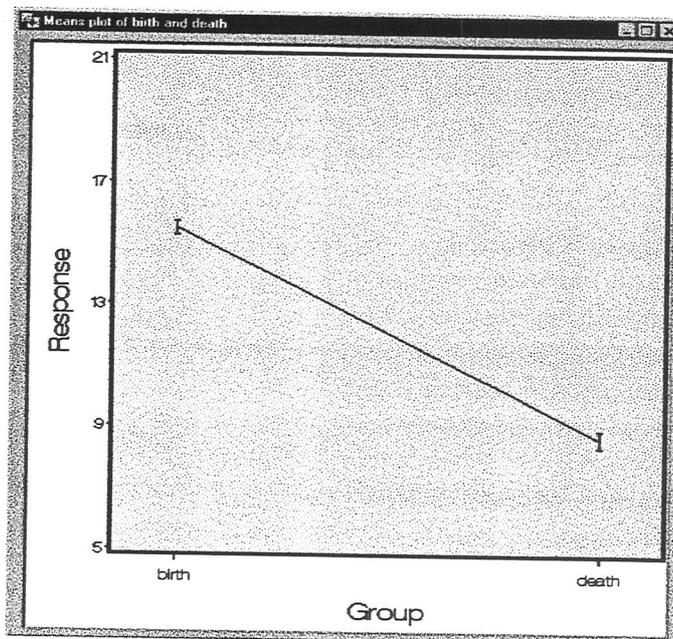


Figura 9.3.11: Teste t para amostras pareadas: Gráfico das Médias.

Exercício 17.

Faça um teste de hipótese para duas amostras pareadas, utilizando o arquivo `pantes_pdepois.xls`.

9.3.4 Teste t para a Diferença Entre Médias de Duas Amostras Independentes

Selecione **Hipotesys Test** do menu **Statistics menu** e então selecione **Two-Sample t-test for Means** para exibir o diálogo do teste t para a diferença entre médias para duas amostras. Você pode testar se as médias de duas populações são iguais ou se elas diferem por uma quantidade especificada. Dados de duas amostras aparecem quando duas amostras independentes são observadas, possivelmente com diferentes tamanhos de amostras. Note que se duas amostras são dependentes, o teste t para duas amostras é inapropriado. Em vez, use a tarefa do teste t para duas amostras pareadas. Na tarefa do teste t para duas amostras, a hipótese de teste será executada sobre a diferença das médias das duas amostras, e você deve especificar um valor para esta diferença. O valor por definição é zero, que resulta em um teste em que as duas médias das populações são iguais. Você deve escolher uma das três hipóteses alternativas.

Resultados do Teste t para Duas Amostras

Os resultados padrões incluem um resumo estatístico para duas amostras bem como duas estatísticas t com os graus de liberdades associados e p-values. A primeira estatística t assume que a variância da população dos dois grupos é igual e é calculada como:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - m}{\sqrt{s^2(1/n_1 + 1/n_2)}}$$

Onde m é a diferença entre as médias especificada na hipótese de nulidade, \bar{x}_1 é a média amostral para a primeira amostra, \bar{x}_2 é a média amostral para a segunda amostra n_1 é o tamanho da primeira amostra e n_2 o tamanho da segunda amostra. A variância para as amostras combinadas é dada por

$$s^2 = [((n_1-1)s_1^2 + (n_2-1)s_2^2)/(n_1+n_2-2)]$$

A segunda estatística é uma estatística t aproximada, e ela deveria ser usada se as variâncias populacionais dos dois grupos são diferentes. Os graus de liberdade para esta estatística são calculados de acordo com a aproximação de Satterthwaite (1946). O p-value é a probabilidade de obter um resultado que é no mínimo tão extremo quanto aquele obtido pelos dados, se a hipótese de nulidade é verdadeira. Seu cálculo depende da estatística t e da hipótese alternativa.

Especificação Amostral

Suas duas amostras de dados independentes podem ser arranjadas em uma das duas maneiras. Na maioria dos casos, os valores a serem analisados serão contidos em uma variável. Se todos os valores estão sobre uma variável no conjunto de dados, então você pode selecionar o botão **One Variable**. Então selecione a variável e coloque-a na caixa **Group**. Por exemplo, se você for comparar a média da altura de homens e mulheres, a variável **Altura** deveria ir na caixa **Dependent**, e a variável **Sexo** deveria ir para a caixa

Group. A variável na caixa **Group** deve ter somente dois valores distintos. Menos comum, duas amostras estarão em diferentes variáveis. Neste caso, selecione o botão **Two Variables** e coloque uma variável na caixa **Group1** e a outra na caixa **Group 2**. Note que freqüentemente quando os dados estão nesta forma, as amostras são dependentes, e o teste t para amostras pareadas seria o teste apropriado para o caso. Por exemplo, se cada observação representa um estudante de escola secundária e as duas variáveis SATM e SATV representam escores de matemática e de avaliação oral, então as duas amostras definidas por aquelas variáveis são dependentes.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las na análise. Um “C” que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Crtl enquanto clica nas variáveis.

Botão Dependent/Group1

Selecione a variável contendo os dados e clique no botão **Dependent** para posicioná-la na caixa. Se seus dados estão em duas variáveis, clique no botão **Group 1** para posicionar a primeira variável nesta caixa.

Botão Dependent/Group 2

Selecione a variável contendo os dados e clique no botão **Dependent** para posicioná-la na caixa. Se seus dados estão em duas variáveis, clique no botão **Group 2** para posicionar a segunda variável nesta caixa.

Botão Remove

O botão **Remove** habilita você a remover variáveis da lista exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove** e a variável será removida da lista. Você pode também remover variáveis da lista dando um duplo clique sobre elas.

Rótulos de Grupos

Depois que você especificar as variáveis, o rótulo dos grupos exibirá a identidade das duas amostras.

Hipóteses de Nulidade e Alternativa

O teste de hipóteses executado na tarefa fornece um método para decidir qual das duas teorias sobre a verdadeira diferença entre as médias das variáveis selecionada é suportada pelos dados. A primeira teoria, conhecida como hipótese de nulidade, representa o status quo. Esta hipótese é preservada a menos que os dados dêem forte evidência contrária. Você deve especificar um valor para a diferença entre as médias para representar o status quo. A hipótese alternativa representa uma mudança de status quo. Você deve especificar uma das três escolhas disponíveis. A primeira escolha é que a diferença entre as

médias não é igual a diferença especificada. Isto é, uma alternativa bilateral. As alternativas unilaterais são que a diferença é maior que ou menor que a diferença especificada. Note que a escolha da hipótese alternativa deveria ser baseada na expectativa do experimento, não sobre os dados.

Hipótese de Nulidade

Especifica um valor para a hipótese de nulidade para a diferença entre as médias. Especificamente testa-se se as médias são iguais, ou seja se a diferença entre elas é nula.

Hipótese Alternativa

Selecione uma hipótese alternativa. As três escolhas são **diferente, maior que, e menor que** o valor da hipótese de nulidade.

Opções do Teste-t para duas amostra

Os seguintes diálogos estão disponíveis para personalizar a sua análise.

- Tests: Produz intervalos de confiança ou análise do poder do teste.
- Plots: Especifica quatro tipos de Gráficos.
- Titles: Especifica os títulos que aparecem nos resultados.
- Variables: Especifica um grupo de variável.

Exercício 18.

Faça um Teste t para a diferença entre duas amostras com os dados do arquivo **laranja.xls**.

9.4 Análise de Variância (ANOVA)

9.4.1 Introdução

A análise de variância é uma técnica para decompor a variação de uma variável resposta contínua (variável dependente). A variável resposta é medida para diferentes níveis de uma ou mais variáveis de classificação (variáveis independentes). A variação na resposta devida as variáveis de classificação é computada e pode-se testar esta variação contra a variação devida ao erro experimental (resíduo) para determinar a significância dos efeitos de classificação.

O menu para a análise de variância é exibido na figura 9.4.1

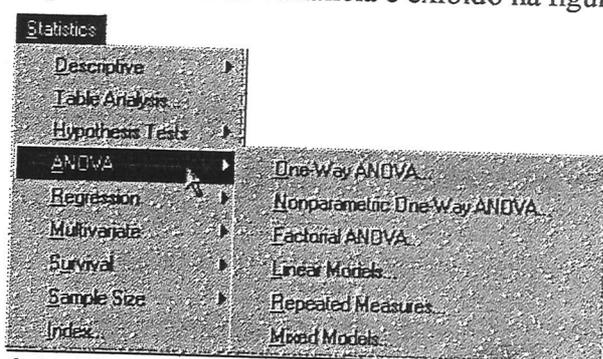


Figura 9.4.1: Menu para Análise de Variância

Exemplo

Conjunto de dados sobre a qualidade do ar

Os dados referem-se a medidas da qualidade do ar, obtidas em um vale industrial. As medidas são tomadas de hora em hora por um período de uma semana.

A primeira variável no conjunto de dados é uma variável do SAS (*datetime*) que contém a data e o tempo do dia no qual a observação foi tomada. O conjunto de dados apresenta duas variáveis adicionais relacionadas ao tempo para a *datetime*, as quais registram o dia da semana e a hora do dia. As variáveis de medida da qualidade do ar são *co* (monóxido de carbono), *o3* (ozônio), *so4* (sulfato), *no* (óxido nitroso), e *pó* (partículas). A variável final fornecida é o vento, a qual fornece a velocidade do vento em milhas marítimas.

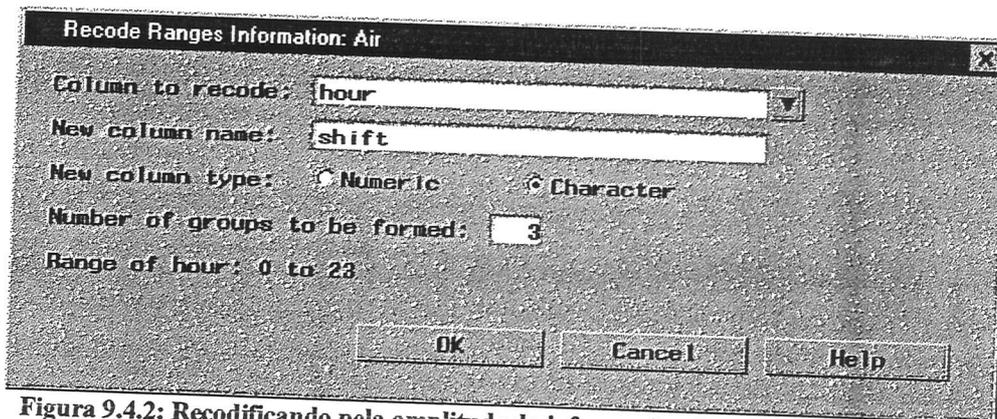
Abrir (*Open*) o arquivo de dados

Os dados fazem parte da Livraria de exemplos do Analyst (Analyst Sample Library). Para acessar o conjunto de dados segue-se as seguintes etapas:

1. Selecione **Tools** → **Sample Data ...**
2. Selecione **Air**.
3. Clique **OK** para criar o conjunto de dados no seu diretório **Sasuser**
4. Selecione **File** → **Open By SAS Name ...**
5. Selecione **Sasuser** da lista de **Libraries**.
6. Selecione **Air** da lista de membros
7. Clique **OK** para abrir o conjunto de dados na sua tabela de dados

Criando uma Nova Variável

Para executar a análise nos exemplos seguintes, você precisa criar uma nova variável para representar os períodos de mudança de turno da fábrica. Uma variável caracter nova, turno, registra a hora variável nos três turnos da fábrica. A figura 9.4.2 mostra o diálogo Recodificação da Amplitude da Informação. Entre com as informações para criar a nova variável como aparece na figura 9.4.2.



Recode Ranges Information: Air

Column to recode: hour

New column name: shift

New column type: Numeric Character

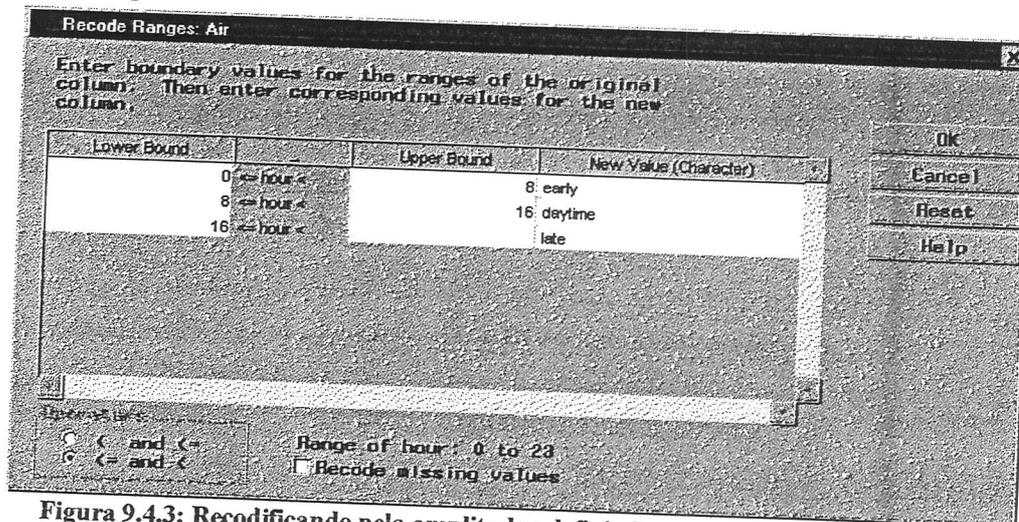
Number of groups to be formed: 3

Range of hour: 0 to 23

OK Cancel Help

Figura 9.4.2: Recodificando pela amplitude da informação: definindo a nova variável

Clique **OK** para exibir o diálogo de recodificação pela amplitude que aparece na figura 9.4.3. Para definir os valores para a nova variável *shift*, entre os valores como aparece na figura 9.4.3.



Recode Ranges: Air

Enter boundary values for the ranges of the original column. Then enter corresponding values for the new column.

Lower Bound	Upper Bound	New Value (Character)
0 ← hour <		8 early
8 ← hour <		16 daytime
16 ← hour <		late

Range of hour: 0 to 23

< and <= <= and <

Recode missing values

OK Cancel Reset Help

Figura 9.4.3: Recodificando pela amplitude : definindo os valores para a nova variável

Os valores para a nova variável *shift* são os seguintes: "early" correspondendo ao período entre 0 e 8, isto é da meia-noite até 8 horas da manhã; "daytime" correspondendo ao período entre 8 e 16, isto é das 8 horas da manhã até 4 horas da tarde; "late" correspondendo ao horário maior ou igual a 16, isto é das 4 horas da tarde até a meia-noite.

9.4.2 Análise de Variância em Classificação Simples (One-Way Analysis of Variance)

Selecione ANOVA do menu **Statistics**, então selecione **One-Way ANOVA** para mostrar o diálogo One-Way ANOVA, para executar uma análise de variância em classificação simples. Esta técnica de análise é utilizada para dados experimentais nos quais há uma variável resposta contínua e uma única variável independente de classificação. A variação total na variável resposta é explicada como a soma das variações dos dados devido aos efeitos da variável de classificação e a variação devido ao erro aleatório. Esta tarefa é apropriada se você tem uma variável dependente contínua e uma única variável de classificação. Na tarefa One-Way ANOVA, você deve especificar uma variável independente e uma ou mais variáveis dependentes.

Resultados One-Way ANOVA

Os resultados da análise incluem informações sobre os níveis da variável independente seguida pela tabela padrão de ANOVA, a qual inclui os Graus de Liberdade, Soma de Quadrados e o Quadrado Médio para o modelo e para o erro aleatório. "R-Square" é o coeficiente de determinação e representa a proporção de variabilidade explicada pelas variáveis independentes. "C.V." é o coeficiente de variação. "Root MSE" é a raiz quadrada do Quadrado Médio para o Erro. A linha da tabela ANOVA que representa a variável independente é impressa novamente. A estatística F é a razão dos Quadrados Médios, e "Pr > F" é o *p*-value associado. *p*-values pequenos indicam que a variável independente é significativa na explicação da variação na variável dependente, isto é, que as médias dos níveis da variável de classificação não são todas as mesmas. Uma resposta adicional pode ser produzida se você especificar testes ou comparações de médias.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las no histograma. Um "C" que procede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionada em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Dependent

Selecione uma ou mais variáveis como variáveis dependentes e coloque-as na lista **Dependent**. Se mais que um variável é selecionada, uma análise univariada separada é executada para cada variável.

Botão Independent

Selecione uma única variável de classificação como a variável independente no modelo.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto a lista de candidatos, na qual todas as variáveis estão no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista por clique duplo nelas.

Opções One-Way ANOVA

Os seguintes diálogos estão disponíveis para personalizar sua análise

- ◆ **Tests:** especifica um número de testes de hipóteses e requer uma análise de poder.
- ◆ **Means:** compara e decompõe as médias dos níveis da variável independente.
- ◆ **Plots:** especifica vários tipos de gráficos.
- ◆ **Titles:** especifica títulos que aparecem nos resultados.
- ◆ **Variables:** especifica um grupo de variáveis.

Exemplo

Considerar o banco de dados sobre a qualidade do ar (Air), descrito na seção anterior. Suponha que você queira comparar o nível de ozônio correspondente para cada um dos três turnos de trabalho da fábrica.

Solicitar One-Way ANOVA

Para solicitar o one-way ANOVA, siga os seguintes passos:

1. Selecione **Statistics ANOVA → One-Way ANOVA ...**
2. Selecione *o3* como variável dependente.
3. Selecione turno (*shift*) como variável independente.

A Figura 9.4.4 define o modelo one-way ANOVA.

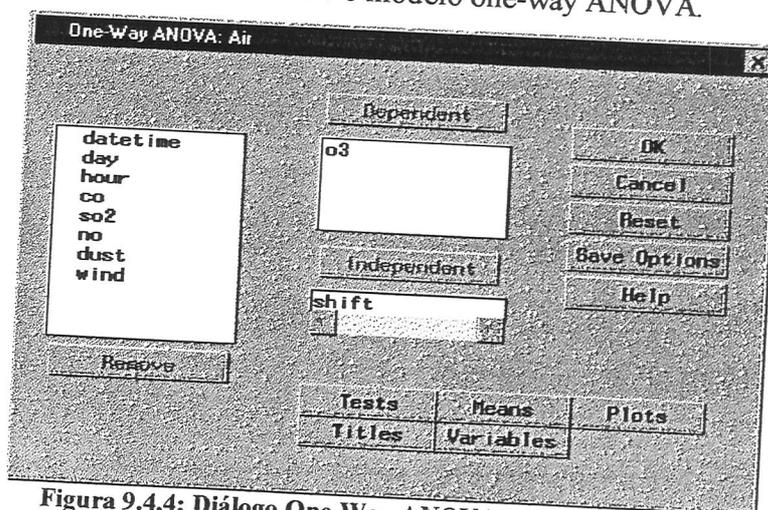


Figura 9.4.4: Diálogo One-Way ANOVA

Solicitar um Teste de Comparação de Médias (*Means Comparison Test*)

A análise de variância executada na tarefa One-Way ANOVA indica se as médias dos grupos são diferentes; não indica qual média em particular é diferente. Para gerar informações mais detalhadas sobre as diferenças entre as médias, siga os seguintes passos:

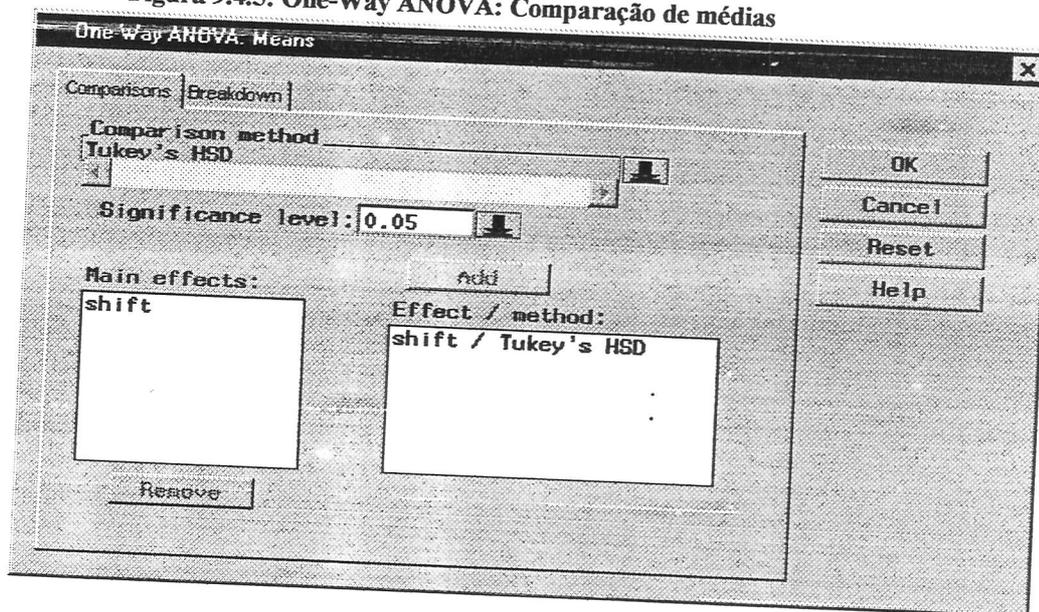
1. Clique no botão **Means** do diálogo principal. A janela resultante mostra o botão **Comparisons**.
2. Clique na seta ao lado da lista **Comparison method**.
3. Selecione **Tukey's HSD**.
4. Aparecerá a variável turno (*shift*) na caixa **Main Effects**.
5. Clique no botão **Add**.

Pode-se clicar na seta próxima a **Significance level** para selecionar o nível de significância, ou pode-se digitar o valor desejado.

6. Clique **OK**.

A Figura 9.4.5 mostra o teste de comparação de médias da amplitude estudentizada de Tukey (HSD), com nível de significância 0.05.

Figura 9.4.5: One-Way ANOVA: Comparação de médias



Solicitar um Gráfico Box-and-Whisker

Para solicitar um gráfico box-and-whisker em adição à análise, siga os seguintes passos:

1. Clique no botão **Plots** do diálogo principal.
2. Selecione **Box-&-whisker plot**.
3. Clique **OK**.

A Figura 9.4.6 mostra o diálogo Plots com **Box-&-whisker plot** selecionado.

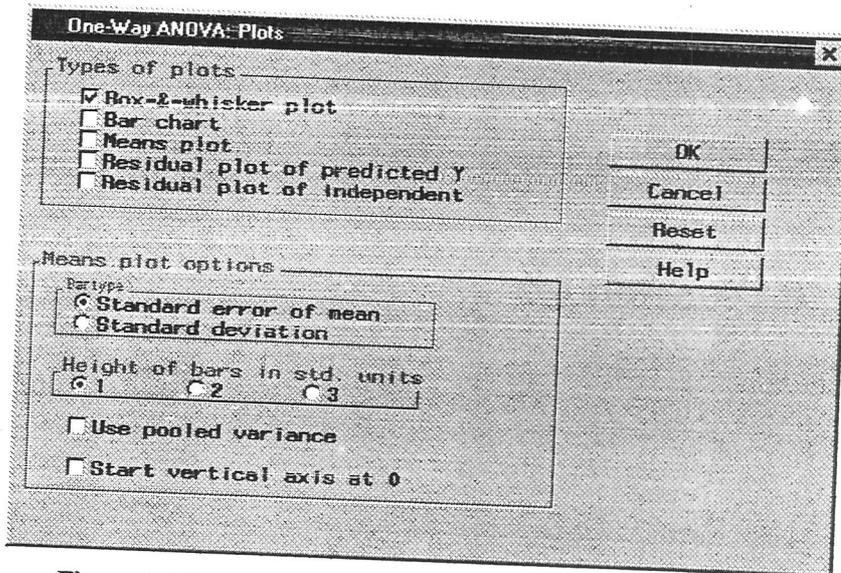


Figura 9.4.6: One-Way ANOVA: Diálogo Plots.

Clique **OK** no diálogo One-Way ANOVA para executar a análise.

Análise dos Resultados

Esta análise testa se a variável independente (*shift*) é um fator significativo na contribuição para a variação dos níveis de ozônio. A Figura 9.4.7 mostra a tabela de análise de variância, com uma estatística F de 31.93 e um p -value associado que é menor que 0.0001. O baixo p -value indica que o modelo explica uma proporção altamente significativa da variação presente na variável dependente.

The ANOVA Procedure

Class Level Information

Class	Levels	Values
shift	3	daytime early late

Number of observations The ANOVA Procedure 168

Dependent Variable: o3 Ozone

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	629.929873	314.964936	31.93	<.0001
Error	165	1627.362727	9.862810		
Corrected Total	167	2257.293599			

R-Square	Coeff Var	Root MSE	o3 Mean
0.279064	73.67665	3.140511	4.262560

Figura 9.4.7: One-Way ANOVA: Resultados da Análise

O valor de R-square, o qual segue a tabela ANOVA Figura 9.4.7, representa a proporção de variação considerada para a variável independente. Aproximadamente 28% da variação no nível de ozônio pode ser responsabilizada pelas diferenças entre turnos.

Informações detalhadas na qual as médias particulares são diferentes está disponível no teste de comparação múltipla, como mostra a Figura 9.4.8. Os resultados da comparação das médias fornece o valor alfa, graus de liberdade e quadrado médio do erro.

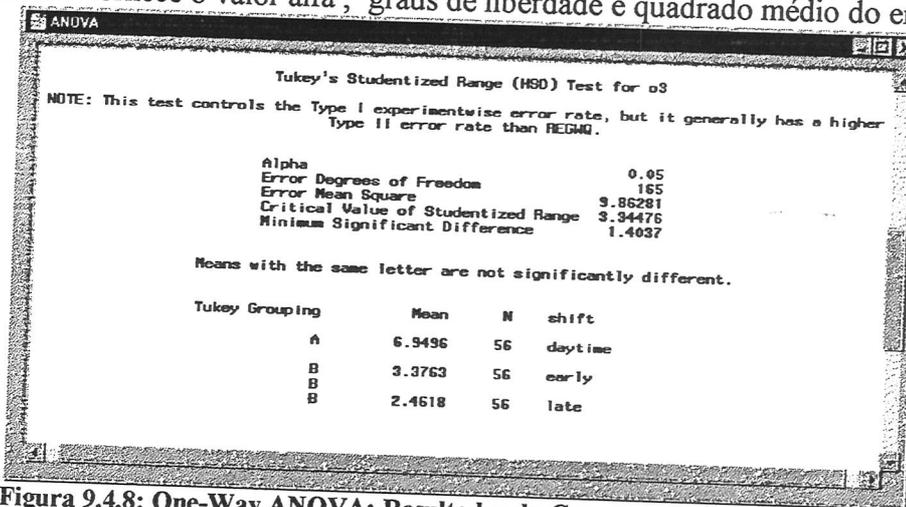


Figura 9.4.8: One-Way ANOVA: Resultados da Comparação Múltipla (Multiple Comparisons Results)

Na tabela "Tukey Grouping", médias com a mesma letra não são significativamente diferentes. A análise mostra que o turno diurno (daytime shift) é associado com níveis de ozônio que são significativamente diferentes dos outros dois turnos. O primeiro e o último turnos não podem ser considerados estatisticamente distintos com base média de níveis de ozônio.

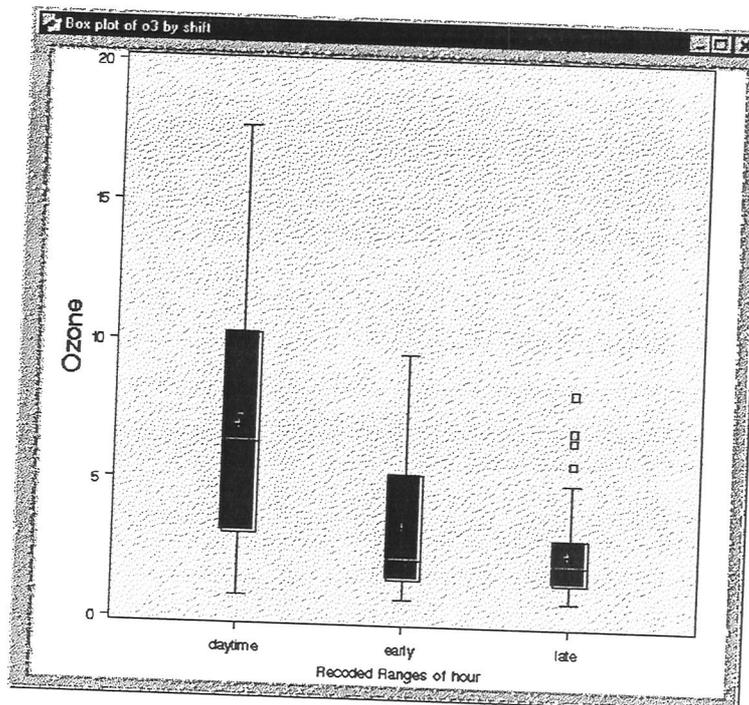


Figura 9.4.9: One-Way ANOVA: Box-and-Whisker Plot

O gráfico box-and-whisker mostrado na Figura 9.4.9 fornece uma visão gráfica da comparação múltipla de resultados. A variação entre os níveis de ozônio talvez seja desigual em cada turno e análises subsequentes podem incluir um teste para

homogeneidade de variâncias, identificando ou não a necessidade de uma transformação da variável resposta o3.

Exercício 19.

Faça uma análise de variância com os dados do arquivo **pulgao.xls**

9.4.3 Análise de Variância para Experimentos Fatoriais (Factorial ANOVA)

Selecione **ANOVA** do menu **Statistics**, então selecione **Factorial ANOVA** para mostrar o diálogo **Factorial ANOVA**, onde você pode executar uma análise de variância numa estrutura fatorial para os tratamentos. Este método de análise é utilizado para dados experimentais nos quais há uma variável resposta contínua e duas ou mais variáveis independentes de classificação dos dados. A variação total nas variáveis resposta é explicada como a soma das variações devido aos efeitos das variáveis de classificação e a variação devido ao erro aleatório. Esta análise é apropriada se você tem múltiplas variáveis de classificação.

Resultado Fatorial ANOVA

Ao resultados da análise incluem informações sobre os níveis das variáveis independentes seguidas pela tabela padrão de ANOVA, a qual inclui os Graus de Liberdade, Soma de Quadrados e o Quadrado Médio para o modelo e para o erro aleatório. “R-Square” é o coeficiente de determinação e representa a proporção de variabilidade explicada pelas variáveis independentes. “C.V.” é o coeficiente de variação. “Root MSE” é a raiz quadrada do Quadrado Médio para o Erro. A estatística F é a razão dos valores dos Quadrados Médios, e “Pr > F” é o *p*-value associado. Um *p*-value pequeno indica que o modelo é significativo para explicar a variação na variável dependente. Abaixo destas informações, a soma de quadrados do modelo é decomposta pelas contribuições dos efeitos no modelo, e os efeitos individuais são analisados pelas suas significâncias. Uma resposta adicional pode ser produzida dependendo das opções que você especifica.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las na análise. Um “C” que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Dependent

Selecione uma ou mais variáveis quantitativas e clique no botão **Dependent** para especificá-las como variáveis dependentes. Se mais de uma variável dependente é selecionada, são executadas análises univariadas separadas para cada variável dependente.

Botão Independent

Selecione uma ou mais variáveis de classificação e clique no botão **Independent** para especificá-las como variáveis independentes no modelo. Por definição o modelo inclui cada variável independente como um efeito principal. Você pode especificar outros modelos no diálogo Model.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto a lista de candidatos, na qual todas as variáveis estão no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista por clique duplo nelas.

Opções Fatorial ANOVA

Os seguintes diálogos estão disponíveis para personalizar sua análise.

- ◆ **Model:** especifica o modelo a ser utilizado na análise.
- ◆ **Tests:** Produz testes de hipóteses, uma análise de poder, ou uma análise de mínimos quadrados ponderados.
- ◆ **Statistics:** especifica quais somas de quadrados produzir.
- ◆ **Means:** requer comparações de médias e médias de mínimos quadrados.
- ◆ **Plots:** especifica gráficos de médias, resíduos, e de influência.
- ◆ **Save Data:** cria um conjunto de dados para resultados.
- ◆ **Titles:** especifica títulos que aparecem nos resultados.
- ◆ **Variables:** especifica variáveis de grupo e frequência.

Exemplo 9.4.3

Considere o banco de dados sobre a qualidade do ar (Air), "The Air Quality Data Set", do exemplo 9.4.1. Suponha que você queira comparar os níveis de ozônio para cada dia da semana e para cada turno de trabalho da fábrica. Você pode definir um modelo fatorial que inclui as duas variáveis de classificação, dia (*day*) e turno (*shift*).

Neste exemplo, um modelo fatorial é especificado, e o gráfico de efeitos de interação é solicitado.

Solicitar a Análise (*Analysis*)

Para solicitar uma análise de variância fatorial, siga os seguintes passos.

1. Clique em **Statistics → ANOVA → Factorial ANOVA ...**
2. Selecione *o3* como variável dependente.
3. Selecione *shift* e *day* como variáveis independentes.

O diálogo Factorial ANOVA resultante é mostrado na Figura 9.4.10.

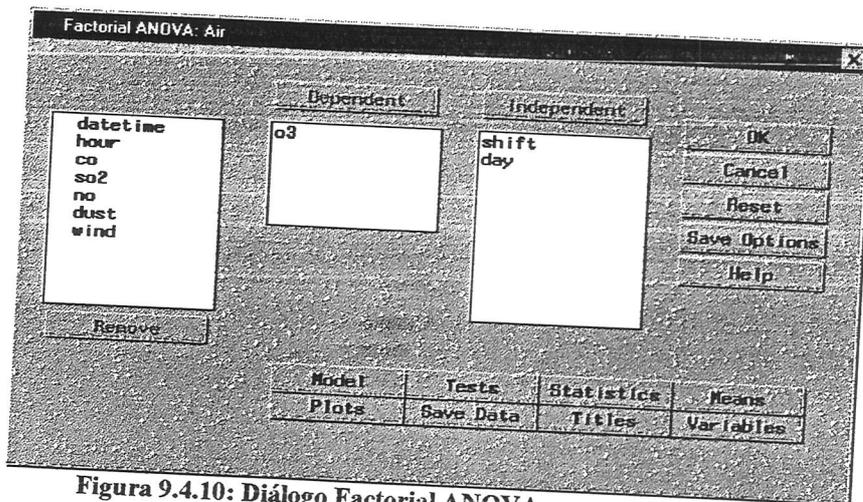


Figura 9.4.10: Diálogo Factorial ANOVA

O modelo ANOVA padrão inclui somente os efeitos principais (isto é, os termos representando turno *-shift*, e dia *-day*). Para incluir um termo de interação, ou especificar outras opções para a sua análise, você pode usar o diálogo disponível na tarefa Factorial ANOVA.

Especificar o Modelo

Para especificar o modelo fatorial, siga os seguintes passos:

1. Clique no botão **Model** do diálogo principal.
2. Selecione as variáveis *shift* e *day* no diálogo de resultados.
3. Clique no botão **Factorial**.
4. Clique **OK**.

A Figura 9.4.11 mostra o diálogo Model com os termos *shift*, *day*, e o termo de interação *shift*day* selecionados como efeitos no modelo.

Você pode construir modelos específicos com os botões **Add**, **Cross**, e **Factorial**, ou você pode selecionar um modelo clicando no botão **Standard Models** e fazendo uma seleção a partir da lista. Desta lista, pode-se solicitar que o seu modelo inclua somente efeitos principais, efeitos bi-direcionais, ou efeitos tri-direcionais.

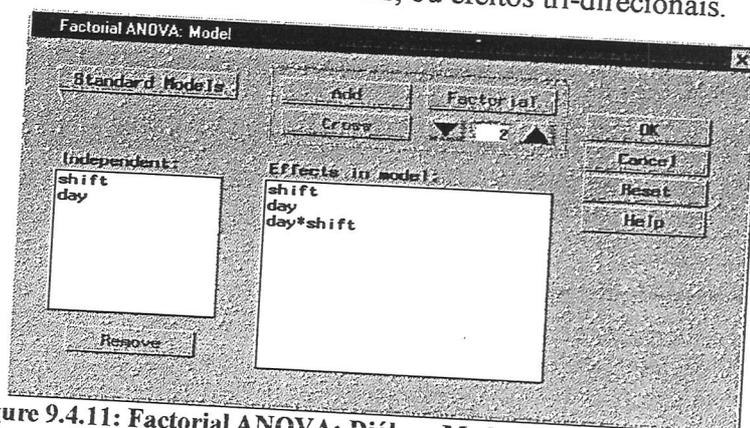


Figure 9.4.11: Factorial ANOVA: Diálogo Modelo.

Solicitar um Gráfico de Médias (*Means Plot*)

Um gráfico de médias mostra um símbolo para as médias observados ou previstos em cada nível de uma variável especificada, com barras verticais estendendo-se para um número especificado de erros padrão. As médias para cada nível de um efeito são unidas com segmentos de retas. Para solicitar um gráfico das médias dependentes, siga os seguintes passos:

1. Clique no botão **Plots** no diálogo principal. A janela de resultados mostra a etiqueta **Means**.
2. Selecione **Plot dependent means for two-way effects**.
Você pode escolher para gráfico as médias observadas ou previstas da variável dependente. Além disso, você pode escolher se as barras verticais deveriam representar um, dois ou três erros padrão.

3. Clique **OK**.

A Figura 9.4.12 solicita um gráfico das médias dependentes observados para os efeitos bi-dimensionais.

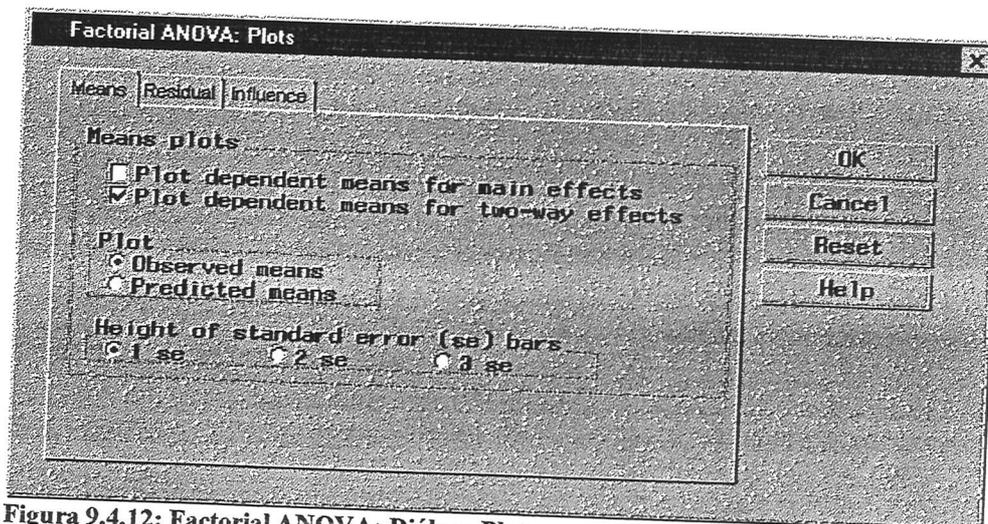


Figura 9.4.12: Factorial ANOVA: Diálogo Plots.

Clique **OK** no diálogo principal para executar a análise.

Análise dos Resultados

A Figura 9.4.13 mostra informações nos níveis de duas variáveis de classificação, *shift* e *day*, seguidas pela tabela ANOVA. A soma de quadrados do modelo é partido nas contribuições separadas dos modelos de efeitos individuais, e testes *F* são fornecidos para cada efeito.

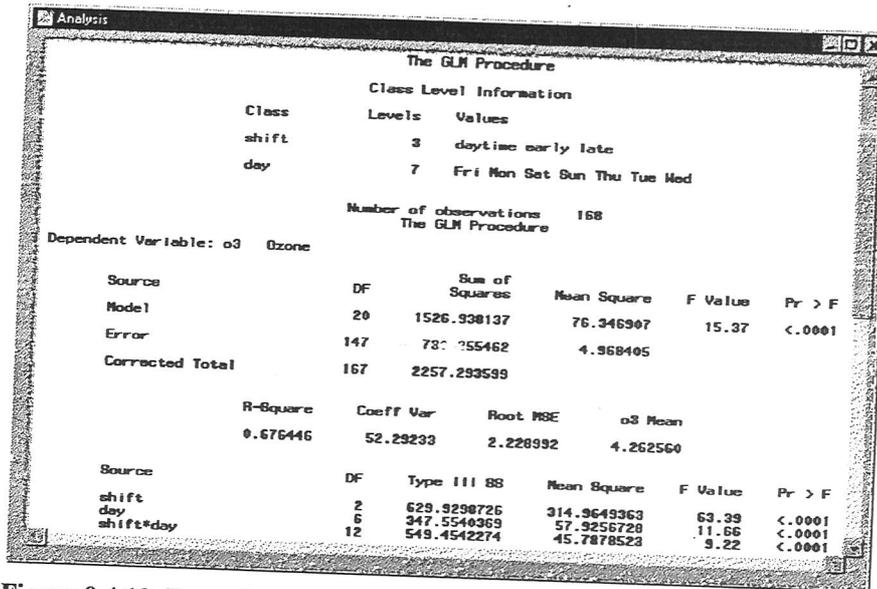


Figura 9.4.13: Factorial ANOVA: Resultados da Análise

A estatística F de 15.37 indica que o modelo de modo geral é altamente significativo (o p -value é menos que 0.0001). Além disso, o R-square de 0.6764 quer dizer que aproximadamente 68% da variação de ozônio pode ser responsabilizada pelo modelo fatorial. A tabela na parte inferior da Figura 9.4.13 mostra o teste de significância para cada termo do modelo. Os efeitos principais e o termo de interação são significantes ao nível $\alpha=0.05$ (cada p -value é muito menor que 0.05).

Na Figura 9.4.14, as três curvas mostram a concentração de ozônio através dos dias da semana. Cada curva representa a relação para um dos três turnos de trabalho.

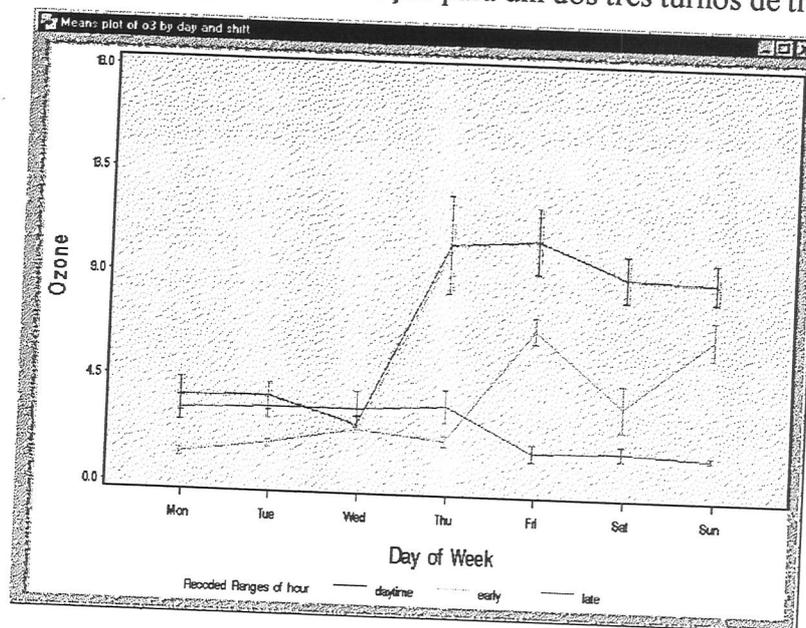


Figura 9.4.14: Factorial ANOVA: Means Plot

O gráfico de médias indica uma relação inversa entre o primeiro e o último turnos. Os níveis de ozônio durante o primeiro turno elevam-se dramaticamente na quinta-feira e

permanecem altos durante todo o fim de semana. Os níveis de ozônio para o último turno, por outro lado, começam a decrescer após a quinta-feira e permanecem baixos durante o fim de semana.

Exercício 20.

Realize uma análise fatorial nos dados do arquivo `anova_fatorial.xls`

9.4.4 Modelos Lineares(Linear Models)

Selecione **ANOVA** do menu *Statistics*, então selecione **Linear Models** para mostrar o diálogo *Linear Models*, onde você pode ajustar modelos lineares em geral com o método de mínimos quadrados. Este método de análise é utilizado para dados experimentais nos quais há uma variável resposta contínua e uma ou mais variáveis independentes de classificação dos dados assim como uma ou mais variáveis independentes quantitativas. A variação total na variável resposta é explicada como a soma das variações devido aos efeitos do modelo e a variação devido ao erro aleatório. Na tarefa *Linear Models*, você deve especificar uma ou mais variáveis dependentes e pelo menos uma variável de classificação ou variável quantitativa.

Resultados de Modelos Lineares

Os resultados da análise incluem informações sobre os níveis das variáveis independentes seguidas pela tabela padrão ANOVA, a qual inclui os Graus de Liberdade, Soma de Quadrados e o Quadrado Médio para o modelo e para o erro aleatório. “R-Square” é o coeficiente de determinação e representa a proporção de variabilidade de responsabilidade das variáveis independentes. “C.V.” é o coeficiente de variação. “Root MSE” é a raiz quadrada do Quadrado Médio para o Erro. A estatística F é a razão dos valores dos Quadrados Médios, e “Pr > F” é o *p*-value associado. Um *p*-value pequeno indica que o modelo é significativo para explicar a variação na variável dependente. Abaixo destas informações, a soma de quadrados do modelo é decomposta pelas contribuições dos efeitos no modelo, e os efeitos individuais são analisados pela sua significância.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las na análise. Um “C” que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Dependent

Selecione uma ou mais variáveis quantitativas como variáveis dependentes e clique no botão **Dependent** para colocá-las na lista *Dependent*.

Botão Class

Selecione uma ou mais variáveis de classificação como variáveis independentes no modelo e clique no botão **Class** para adicionar para a lista *Class*. Personalize o modelo no

diálogo Model. Por definição o modelo inclui o efeito principal de cada variável independente de classificação e quantitativa.

Botão Quantitative

Selecione uma ou mais variáveis quantitativas como variáveis independentes no modelo e clique no botão **Quantitative** para adicioná-las a lista Quantitative. Personalize o modelo no diálogo Model. Por definição o modelo inclui o efeito principal de cada variável independente de classificação e quantitativa.

Botão Remove

O botão **Remove** permite que você remova variáveis de qualquer lista exceto a lista de candidatos, na qual todas as variáveis estão no conjunto de dados. Selecione a variável ou variáveis, então clique no botão **Remove**, e as variáveis serão removidas da lista. Você também pode remover variáveis da lista por clique duplo nelas.

Opções de Modelos Lineares

Os seguintes diálogos estão disponíveis para personalizar sua análise.

- ◆ **Model**: especifica o modelo a ser utilizado na análise.
- ◆ **Tests**: produz testes de hipóteses, uma análise de poder, ou uma análise de mínimos quadrados ponderados.
- ◆ **Statistics**: especifica quais somas de quadrados produzir.
- ◆ **Means**: requer comparações de médias e médias de mínimos quadrados.
- ◆ **Predictions**: prediz respostas baseadas na análise.
- ◆ **Plots**: especifica gráficos de médias, de valores preditos, de resíduos, e de influência.
- ◆ **Save Data**: cria um conjunto de dados para resultado.
- ◆ **Titles**: especifica títulos que aparecem nos resultados.
- ◆ **Variables**: especifica variáveis de grupo e frequência.

Exemplo

O banco de dados Air, descrito anteriormente, inclui medidas quantitativas; por exemplo, a variável *wind* representa a velocidade do vento, em nós. Suponha que você queira modelar níveis de ozônio usando as variáveis *day* (dia da semana), *shift* (turno de trabalho das fábricas), e *wind* (velocidade do vento.). Suponha que você também queira seu modelo para incluir a interação entre as variáveis *day* e *shift*. Quer dizer, você quer executar uma análise bidirecional simples da covariância da desigualdade do declive. O exemplo a seguir prepara este modelo linear e adicionalmente solicita uma análise retrospectiva (retrospective power analysis) e um gráfico dos valores observados versus os valores previstos.

Solicitar o Modelo de Análise Linear (*Linear Models Analysis*)

Para solicitar o modelo de análise linear, siga os seguintes passos:

1. Selecione **Statistics** → **ANOVA** → **Linear Models** ...
2. Selecione *o3* como a variável dependente.

3. Selecione *shift* e *day* como variáveis de classificação.
4. Selecione *wind* como a variável quantitativa

A Figura 9.4.16 mostra o diálogo Linear Models.

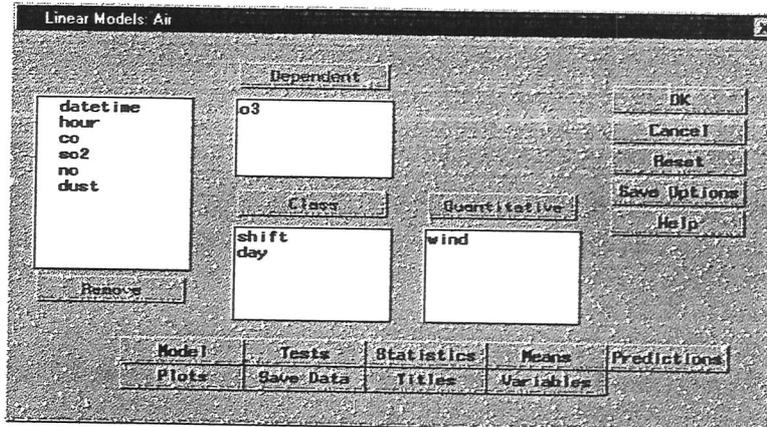


Figura 9.4.16: Diálogo Linear Models

Por definição, a análise linear de modelos inclui somente os efeitos principais especificados no diálogo principal: não estão incluídos termos de interação.

Especificar um Termo de Interação (*Interaction Term*) no Modelo

Para incluir o termo de interação *shift*day* no modelo, siga os seguintes passos:

1. Clique no botão **Model** no diálogo principal.
2. Selecione as variáveis *shift* e *day*.
3. Clique no botão **Cross**.
4. Clique **OK**.

Você pode construir modelos específicos com os botões **Add**, **Cross**, e **Factorial**, ou você pode selecionar um modelo clicando no botão **Standard Models** e fazendo uma seleção a partir da lista.

A Figura 9.4.17 mostra o diálogo Model com os termos *shift* e *day* e o termo de interação *shift*day* selecionados como efeitos no modelo.

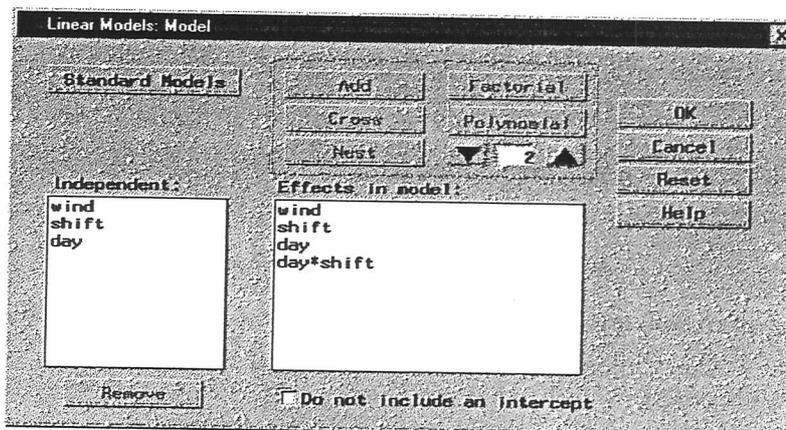


Figure 9.4.17: Linear Models: Diálogo Model.

Solicitar uma Análise de Poder (*Power Analysis*)

O poder de um teste é a probabilidade de corretamente rejeitar as hipóteses de nulidade da não diferença. Isso depende do tamanho da amostra bem como da precisa diferença especificada na hipótese alternativa. Em condições ideais, você considera o poder antes de coletar os dados para assegurar que você vai coletar dados suficientes para detectar uma diferença. De qualquer forma, uma vez que você tiver com os dados coletados, você pode executar uma análise retrospectiva de poder, para determinar quanto mais dados são necessários para detectar a diferença observada. Para executar uma análise retrospectiva de poder com a aplicação do Analyst, siga os seguinte passos:

1. Clique no botão **Tests** no diálogo principal.
2. Clique na placa **Power Analysis**.
3. Selecione **Perform power analysis**.

Para solicitar cálculos do poder para testes executados com vários valores de alfa, você pode entrar com os valores, separados por um espaço, na caixa intitulada **Alphas**. Você pode solicitar a análise do poder para tamanhos de amostra adicionais na caixa **Sample sizes**. Você pode entrar com um ou mais valores específicos para o tamanho da amostra, ou especificar uma série de tamanhos de amostra nas caixas intituladas **From**,

To, e **By**:

4. Clique **OK**.

A Figura 9.4.17 mostra a placa **Power Analysis**, que pede uma análise retrospectiva de poder com um alpha, ou nível de significância, de 0.05.

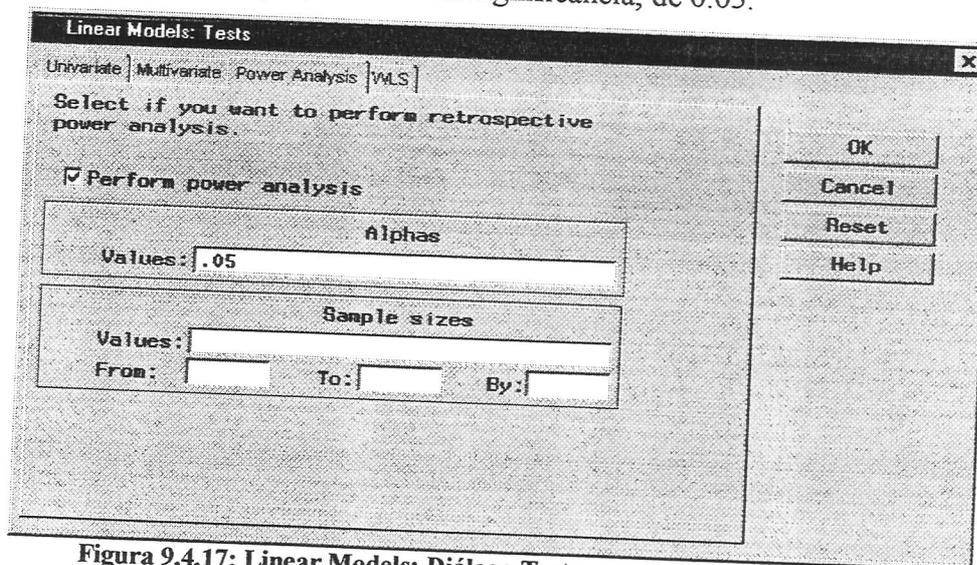


Figura 9.4.17: Linear Models: Diálogo Tests.

Solicitar um Gráfico de Dispersão (*Scatter Plot*)

Para solicitar um gráfico de dispersão dos valores preditos versus os valores observados, siga os seguintes passos:

1. Clique no botão **Plots** do diálogo principal.
2. Clique na placa **Predicted**.
3. Selecione **Plot observed vs predicted**.
4. Clique **OK**.

A Figura 9.4.18 mostra a placa **Predicted** no diálogo Plots.

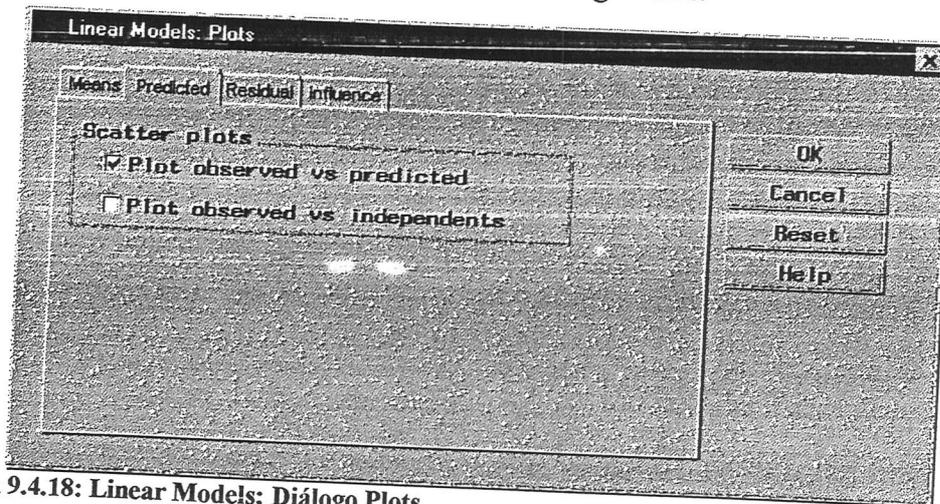


Figura 9.4.18: Linear Models: Diálogo Plots.

Clique **OK** no Diálogo Linear Models para executar a análise.

Análise dos Resultados

A saída da análise inclui informações sobre os níveis das variáveis independentes, seguidos pela tabela ANOVA.

A Figura 9.4.19 mostra a tabela de análise de variância, com uma estatística F de 19.44 e um p -value associado menor que 0.0001. Um p -value tão baixo indica que o modelo explica uma proporção altamente significativa de variação na variável dependente.

O valor R-square representa a proporção da variabilidade de responsabilidade das variáveis independentes. Nesta análise, aproximadamente 74% da variação do nível de ozônio pode ser atribuída ao modelo (isto é, pelas diferenças das médias no dia e turno, em combinação com uma dependência linear na velocidade do vento *-wind*).

Analysis

The GLM Procedure

Class Level Information	
Class	Levels Values
shift	3 daytime early late
day	7 Fri Mon Sat Sun Thu Tue Wed

Number of observations The GLM Procedure 168

Dependent Variable: o3 Ozone

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	21	1662.604496	79.171643	19.44	<.0001
Error	146	594.689103	4.073213		
Corrected Total	167	2257.293599			

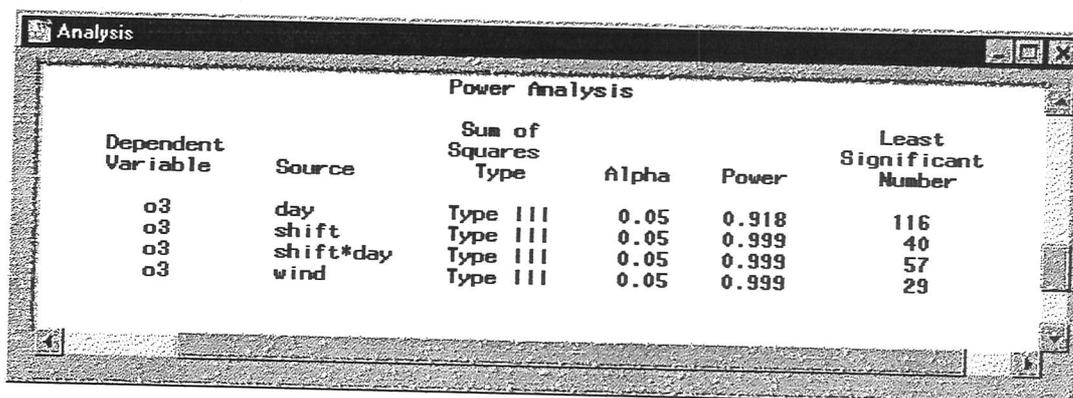
	R-Square	Coeff Var	Root MSE	o3 Mean
	0.736548	47.34761	2.018220	4.262560

Source	DF	Type III SS	Mean Square	F Value	Pr > F
wind	1	135.6663592	135.6663592	33.31	<.0001
shift	2	122.4728208	61.2364104	15.03	<.0001
day	6	78.4551533	13.0758589	3.21	0.0054
shift*day	12	295.5940454	24.6328371	6.05	<.0001

Figura 9.4.19: Linear Models: Resultados ANOVA.

A última tabela mostrada na Figura 9.4.19 a soma de quadrados do modelo é desdobrada em contribuições separadas para cada efeito do modelo e testes para a significância de cada efeito são apresentados. Os efeitos principais e o termo de interação são significantes ao nível $\alpha=0.05$ (isto é, cada p -value é menor que 0.05).

A figura 9.4.20 mostra a análise retrospectiva do poder. O poder observado é determinado para cada efeito no modelo linear.



Dependent Variable	Source	Sum of Squares Type	Alpha	Power	Least Significant Number
o3	day	Type III	0.05	0.918	116
o3	shift	Type III	0.05	0.999	40
o3	shift*day	Type III	0.05	0.999	57
o3	wind	Type III	0.05	0.999	29

Figura 9.4.20 Modelos Lineares: Análise de Poder

A coluna chamada Least Significant Number na Figura 9.4.20 mostra o menor número de observações exigido para determinar que o efeito é significativo para o valor α dado.

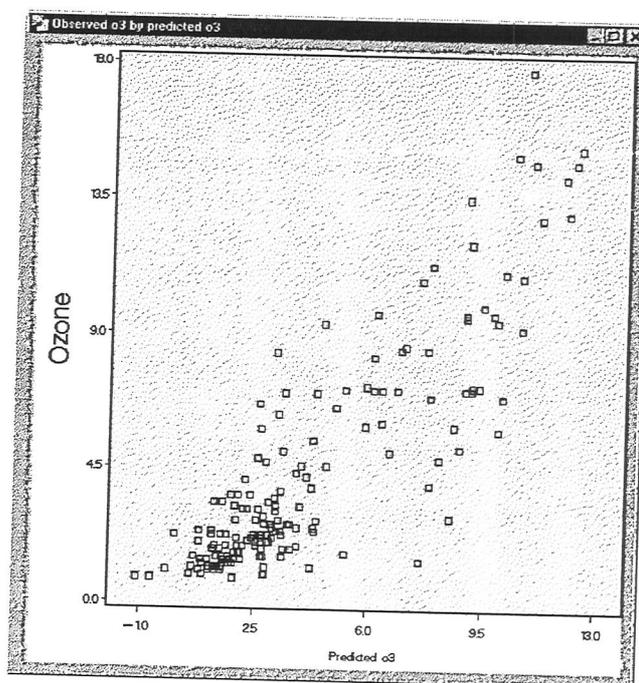


Figura 9.4.21: Linear Models: Níveis de Ozônio Observados versus Valores Preditos.

A Figura 9.4.22 mostra o gráfico dos valores observados versus os valores previstos pelo modelo. Se o modelo predizer os valores observados perfeitamente, os pontos no gráfico configurarão uma linha reta com uma inclinação de 1. Este gráfico indica uma previsão razoável.

Exercício 21

Faça uma análise de variância para modelos lineares com o arquivo **maca.xls**

9.5 Regressão

9.5.1 Introdução

Regressão é uma técnica estatística utilizada para investigar a relação entre uma variável dependente, também chamada variável resposta, e uma ou mais variáveis explicativas, também chamadas preditoras ou variáveis independentes. Na regressão linear a variável dependente é modelada como uma função linear das variáveis quantitativas independentes. Por exemplo, pode-se escrever a equação de regressão linear simples como $Y = b_0 + b_1 X$, onde Y representa a variável dependente, X é a variável explicativa, b_0 e b_1 são os coeficientes da equação, sendo b_0 o intercepto e b_1 o coeficiente de regressão linear.

O menu para a análise de regressão aparece na figura 9.5.1

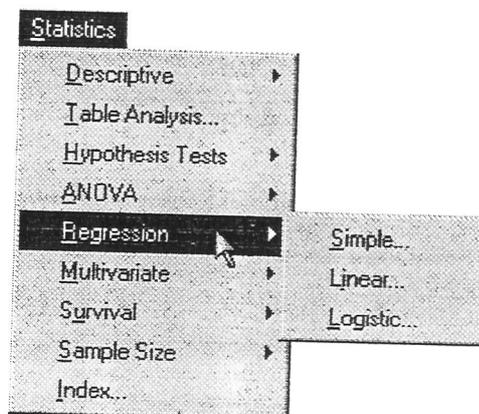


Figura 9.5.1 - Menu para a Análise de Regressão

9.5.2 Regressão Linear Simples

Selecione **Regression** no Menu **Statistics**, então selecione **Simple** para exibir o diálogo da Regressão Linear Simples. Use esta análise quando você tem uma única variável dependente e uma única variável independente quantitativa. A análise fornece o ajuste linear de mínimos quadrados para os dados.

Resultados da Regressão Linear Simples

A saída padrão inclui uma tabela de análise de variância, assim como a raiz quadrada do quadrado médio do erro ("Root MSE"), a média da variável dependente, o coeficiente de variação, o valor do R^2 (coeficiente de determinação), assim como seu valor ajustado ("adjusted R-square"). Na tabela de ANOVA, um p-value baixo (registrado sob o título "Prob>F") indica que o modelo explica uma parte significativa da variação dos dados. A qualidade das estimativas dos parâmetros depende das suposições feitas sobre os dados. Cada p-value registrado sob a denominação de "Prob > |t|" representa a significância para o teste se o parâmetro é significativamente diferente de zero.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las na análise. Um “C” que precede a variável indica que ela é uma variável não numérica (caracter). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais você pode pressionar a tecla Ctrl enquanto clica nas variáveis.

Botão Dependent

Selecione uma variável quantitativa como uma variável resposta e clique no botão **Dependent**.

Botão Explanatory

Selecione uma variável quantitativa como uma variável independente e clique no botão **Explanatory** (explicativa).

Botão Remove

O botão **Remove** permite você remover variáveis de qualquer lista, exceto a lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecionada a variável ou variáveis, clique no botão **Remove**, e a variável desejada será removida da lista. Você também pode remover variáveis da lista dando um clique duplo nelas.

Modelo

Selecione **Linear** para a regressão linear simples usual. Selecione **Quadratic** ou **Cubic** para ajustar regressões polinomiais, de 2º ou 3º grau, ou seja a quadrática e a cúbica.

Opções da Regressão Linear Simples

O diálogo seguinte está disponível para adaptar as análises a seu gosto pessoal:

- ◆ **Testes:** solicitação de uma análise de poder.
- ◆ **Estatísticas:** solicitação adicional de informações das estimativas dos parâmetros.
- ◆ **Preditos:** respostas preditas baseadas na análise.
- ◆ **Gráficos:** especificação de vários tipos de gráficos.
- ◆ **Salvar dados:** criar conjunto de dados para os resultados.
- ◆ **Títulos:** especificar títulos que aparecem na saída.
- ◆ **Variáveis:** especificar variáveis de grupos e frequência.

Exemplo 9.5.1

O conjunto de dados analisado neste exemplo é chamado Houses, e contém as características de quinze casas em venda. O conjunto de dados contém as variáveis que seguem.

Estilo

categoria de estilo (fazenda, dois níveis, condomínio, dois andares)

Área

área em pés quadrados

Quartos

número de quartos

Banheiros

número de banheiros

Rua

nome da rua na qual está localizada a casa

Preço

preço pedido pela casa

A tarefa inclui a execução de uma análise de regressão simples para prever a variável preço da variável explicativa em função da área.

Abrindo o Conjunto de Dados Houses

Os dados são fornecidos pela Analyst Sample Library. Para abrir o conjunto de dados Houses, siga os seguintes passos:

1. Selecione **Tools Sample Data ...**
2. Selecione Houses.
3. Clique **OK** para criar o conjunto de dados amostrais no seu diretório Sasuser.
4. Selecione **File** → **Open By SAS Name ...**
5. Selecione Sasuser da lista de **Libraries**.
6. Selecione Houses da lista de membros.
7. Clique **OK** para trazer o conjunto de dados Houses para a tabela de dados.

Solicitando a Análise de Regressão Simples

Para solicitar a análise de regressão simples, siga estes passos:

1. Selecione **Statistics** → **Regression** → **Simple ...**
2. Selecione preço da lista de candidatos como a variável Dependente.
3. Selecione área da lista de candidatos como a variável Explicativa.

A figura 9.5.2 mostra o diálogo resultante.

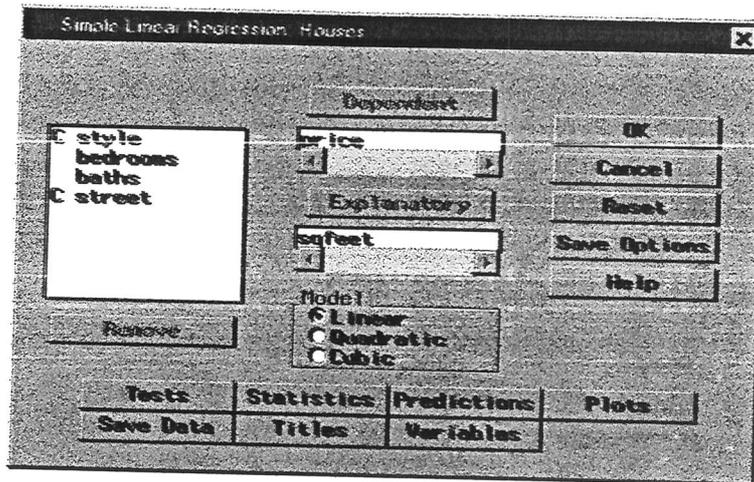


Figura 9.5.2: Diálogo Regressão Linear Simples

O modelo definido nesta análise é

$$\text{Preço} = b_0 + b_1 \text{área}$$

Se você selecionar **Quadratic** ou **Cubic** na caixa **Model**, o modelo respectivo é

$$\text{Preço} = b_0 + b_1 \text{área} + b_2 \text{área}^2$$

ou

$$\text{Preço} = b_0 + b_1 \text{área} + b_2 \text{área}^2 + b_3 \text{área}^3$$

Por definição, a análise ajusta o modelo de regressão simples.

Solicitando um Gráfico de Dispersão dos Dados

Para solicitar um gráfico dos valores observados versus valores independentes, siga estes passos.

1. Clique no botão **Plots**.
2. Selecione **Plot observed vs independent**.

Você pode adicionar limites de confiança para a média da variável independente selecionando **Confidence limits**, ou você pode produzir limites preditos de 95% para previsões individuais.

3. Clique **OK**

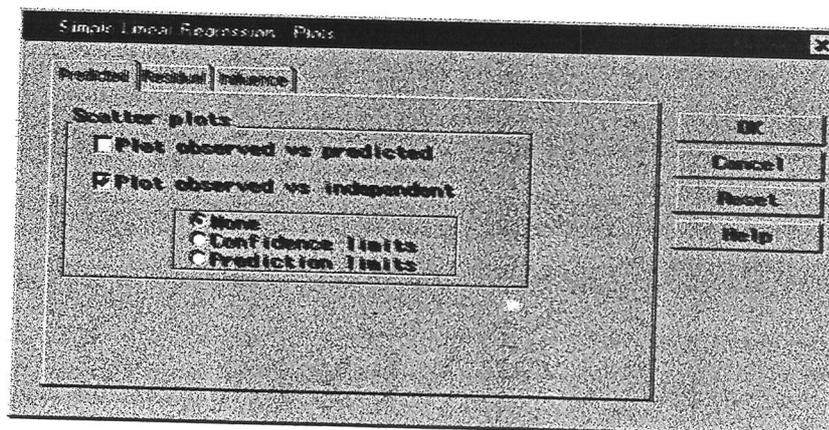


Figura 9.5.3: Regressão Linear Simples: Diálogo Plots

Clique **OK** no diálogo Simple Linear Regression para executar a análise.

Examinando os Resultados

Os resultados são vistos na figura 9.5.4. A tabela ANOVA é vista nos resultados, seguida pela tabela das estimativas dos parâmetros. O ajuste de mínimos quadrados é $\text{Preço} = -14982 + 67.52 \times \text{área}$

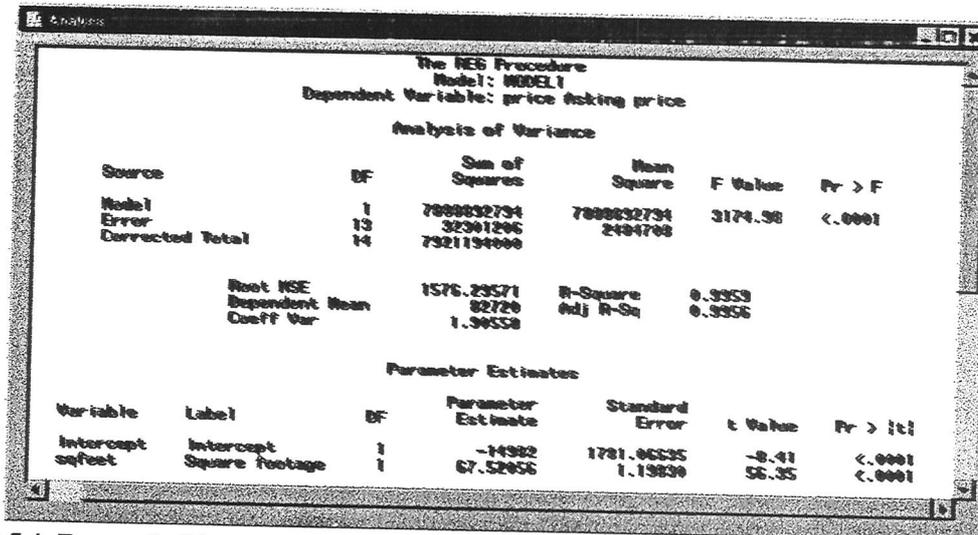


Figura 9.5.4: Regressão Linear Simples: Resultados

Os p -values pequenos listados na coluna $\text{Pr} > |t|$ indicam que ambas as estimativas dos parâmetros são significativamente diferentes de zero.

O gráfico das variáveis resposta e independente é mostrado na figura 9.5.5. O gráfico inclui a linha de regressão ajustada.

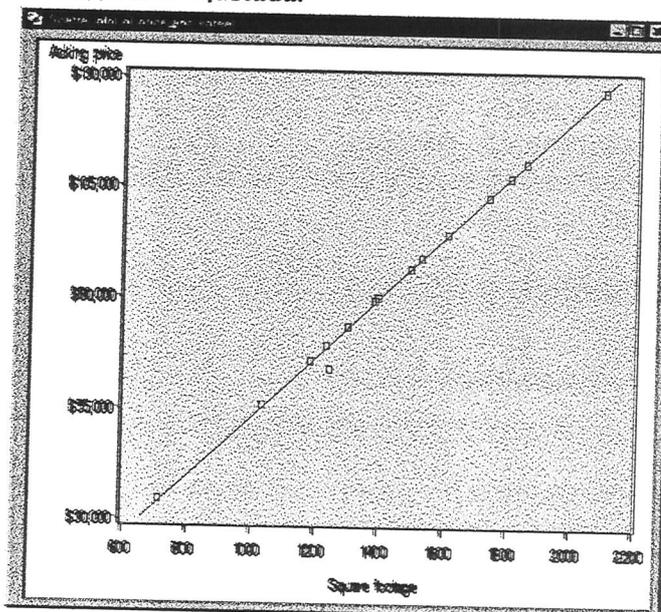


Figura 9.5.5: Regressão Linear Simples: Gráfico de Dispersão com Linha de Regressão

Exercício 22

Faça uma regressão linear simples com os dados do arquivo **mandioca.xls**.

9.5.3 Regressão Linear

A análise de regressão linear múltipla aplica-se quando tem-se mais do que uma variável explicativa a ser considerada no modelo a ser ajustado. A equação de regressão linear múltipla para um modelo com p variáveis explicativas é dada por

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_pX_p,$$

onde Y é a resposta ou variável dependente; X_1, X_2, \dots, X_p representam as p variáveis explicativas ou independentes; b_0 é o intercepto e b_1, b_2, \dots, b_p são os coeficientes de regressão parciais.

Selecione **Regression** no Menu **Statistics** e então selecione **Linear** para exibir o Diálogo da Regressão Linear. Use esta análise para executar a regressão linear para múltiplas variáveis dependentes e independentes quantitativas. A análise fornece o ajuste linear por mínimos quadrados dos dados.

Resultados da Regressão Linear

A saída padrão inclui uma tabela de análise de variância, bem como a raiz quadrada do quadrado médio do erro ("Root MSE"), a média das variáveis dependentes, o coeficiente de variação, o valor do R^2 e o R^2 ajustado. Na tabela de ANOVA, um pequeno p -value (listado abaixo do título "Prob>F") indica que o modelo explica uma fração significativa da variação dos dados. A qualidade das estimativas dos parâmetros depende das suposições assumidas para os dados. Cada p -value listado abaixo do título "Prob > |t|" representa a significância para testar se o parâmetro é significativamente diferente de zero.

Lista de candidatos

A lista de candidatos exibe todas as variáveis do conjunto de dados. Você pode selecionar variáveis a partir da lista para exibi-las na análise. Um "C" que precede a variável indica que ela é uma variável não numérica (character). Note que selecionando variáveis não contíguas da lista de candidatos somente a última permanece selecionada. Se você quer selecionar variáveis posicionadas em muitos locais você pode pressionar a tecla Crtl enquanto clica nas variáveis.

Botão Dependent

Selecione uma ou mais variáveis quantitativas como variáveis dependentes e clique no Botão **Dependent** para adicionar elas na lista de dependentes.

Botão Explanatory

Selecione uma ou mais variáveis quantitativas como variáveis explicativas e clique no botão **Explanatory** para adicionar elas na lista de explicativas.

Remove Button

O botão **Remove** permite você remover variáveis de qualquer lista, exceto da lista de candidatos, que lista todas as variáveis do conjunto de dados. Selecionada a variável ou variáveis, clique no botão **Remove**, e a variável desejada será removida da lista. Você também pode remover variáveis da lista dando um clique duplo nelas.

Opções da Regressão Linear

O diálogo seguinte está disponível para adaptar as análises a seu gosto pessoal:

- ◆ **Model:** especifica um método de seleção de variáveis para o modelo.
- ◆ **Testes:** produz uma análise do poder ou uma análise de mínimos quadrados ponderados.
- ◆ **Estatísticas:** solicitação de estatísticas adicionais e alguns testes de hipóteses.
- ◆ **Preditos:** prediz respostas baseadas na análise.
- ◆ **Gráficos:** especifica gráficos de preditos, resíduos e de influência.
- ◆ **Salvar dados:** cria conjunto de dados para resultados.
- ◆ **Títulos:** especifica títulos que aparecem na saída.
- ◆ **Variáveis:** especifica variáveis de grupos e frequência.

Exemplo 9.5.2

Consideremos o conjunto de dados fitness do exemplo 9.1.4 e suponha que você gostaria de modelar a aptidão aeróbica de uma pessoa conforme a medida de consumo de oxigênio. O objetivo do estudo é prever o preparo físico como medida de consumo de oxigênio. Assim, a variável dependente para a análise é a variável oxigênio. Você pode escolher qualquer das outras variáveis quantitativas (idade, peso, tempo de corrida, batimento cardíaco em descanso, batimento cardíaco enquanto corre, máximo batimento cardíaco durante a corrida) como suas variáveis explicativas.

Suponha que estudos prévios indicam que o consumo de oxigênio é dependente da idade, do tempo tomado para correr 1.5 milhas e o batimento cardíaco durante a corrida. Assim, em ordem para prever o consumo de oxigênio, você estima os parâmetros na seguinte equação de regressão linear múltipla:

$$\text{oxygen} = b_0 + b_1\text{age} + b_2\text{runtime} + b_3\text{runpulse}$$

Esta tarefa inclui execução de uma análise de regressão linear para prever a variável oxigênio em função das variáveis explicativas idade, tempo de corrida e runpulse (batimento cardíaco enquanto corre). Adicionalmente, solicitando-se intervalos de confiança para as estimativas, uma análise de colinearidade e um gráfico de dispersão dos resíduos.

Abrindo o Conjunto de Dados Fitness

Os dados são fornecidos pela Analyst Sample Library. Para acessar este conjunto de dados, siga estes passos:

1. Selecione **Tools** → **Sample Data ...**
2. Selecione **Fitness**.
3. Clique **OK** para criar o conjunto de dados amostrais no seu diretório Sansuser.
4. Selecione **File** → **Open By SAS Name ...**
5. Selecione **Sasuser** da lista de **Libraries**.
6. Selecione **Fitness** da lista de membros.

7. Clique **OK** para trazer o conjunto de dados Fitness para a tabela de dados.

Solicitando a Análise de Regressão Linear

Para especificar a análise, siga estes passos:

1. Selecione **Statistics** → **Regression** → **Linear ...**
2. Selecione a variável oxigênio da lista de candidatos como a variável dependente.
3. Selecione as variáveis idade, tempo de corrida, e runpulse como as variáveis explicativas.

A figura 9.5.6 mostra a tarefa Regression Linear resultante.

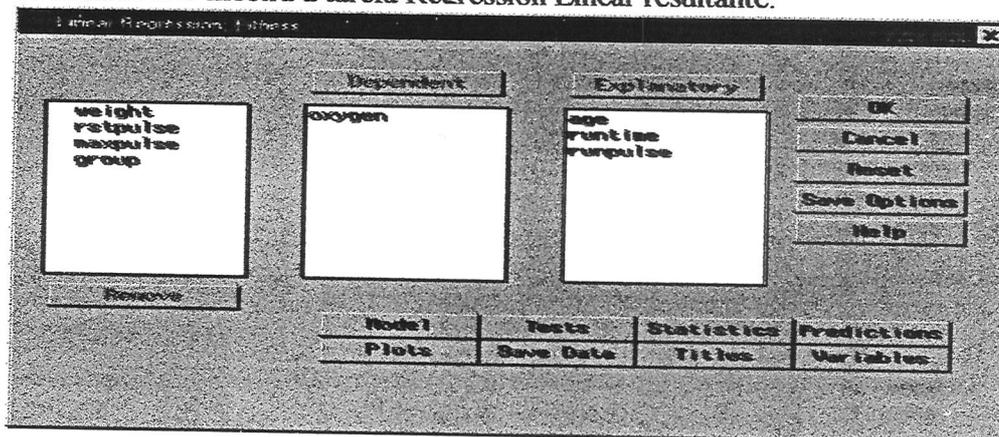


Figura 9.5.6: Diálogo Linear Regression

Por definição, a análise ajusta o modelo de regressão linear.

Solicitando Estatísticas Adicionais

Você pode solicitar várias estatísticas adicionais para a sua análise no diálogo **Statistics**.

Para solicitar que limites de confiança sejam computados, siga estes passos:

1. Clique no botão **Statistics**.
2. Na placa **Statistics**, selecione **Confidence limits for estimates**.

A figura 9.5.7 mostra a placa **Statistics** no diálogo **Statistics**.

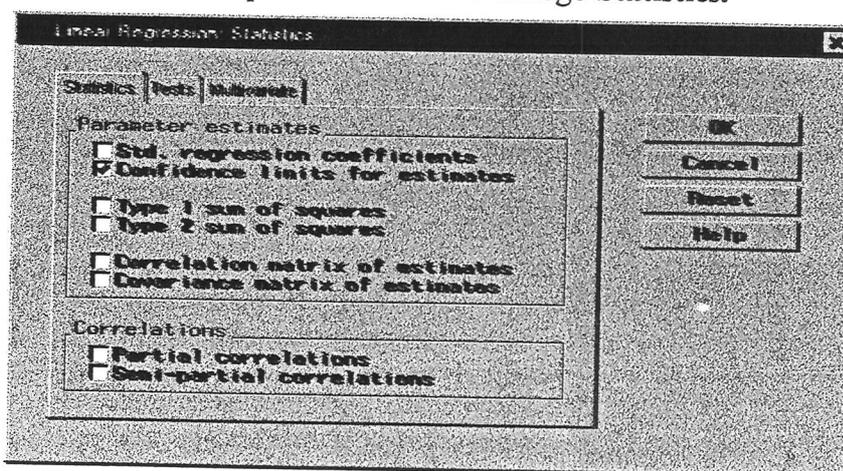


Figura 9.5.7: Regressão Linear: Diálogo Statistics, Placa Statistics

Para solicitar uma análise de colinearidade, siga os seguintes passos:

1. Clique na placa **Tests** no diálogo **Statistics**.
2. Selecione **Collinearity analysis**.
3. Clique **OK**.

O diálogo na figura 9.5.8 requer uma análise de colinearidade para avaliar a estrutura de dependências entre as variáveis explicativas.

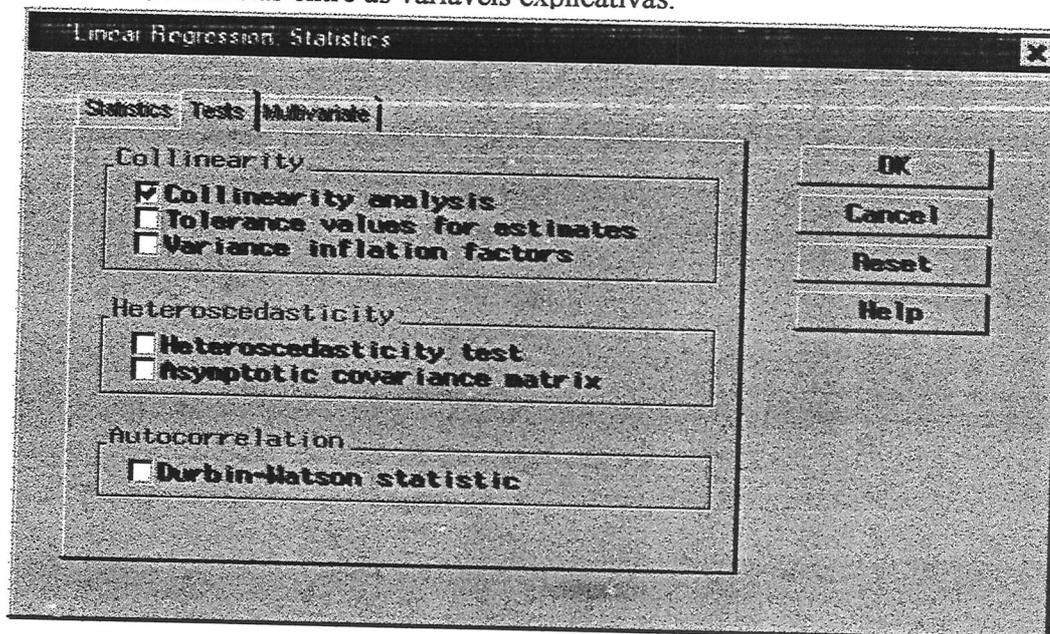


Figura 9.5.8: Regressão Linear: Diálogo Statistics, Placa Tests

Solicitando um Gráfico de Dispersão dos Resíduos

Para solicitar um gráfico dos resíduos estudentizados versus valores preditos, siga estes passos:

1. No diálogo principal Regressão Linear, clique no botão **Plots**.
2. Clique na placa **Residual**.
3. Selecione **Plot residuals vs variables**.
4. Na caixa rotulada **Residuals**, verifique a seleção **Studentized**.
5. Na caixa rotulada **Variables**, verifique a seleção **Predicted Y**.
6. Clique **OK**.

A figura 9.5.9 mostra a placa **Residual**.

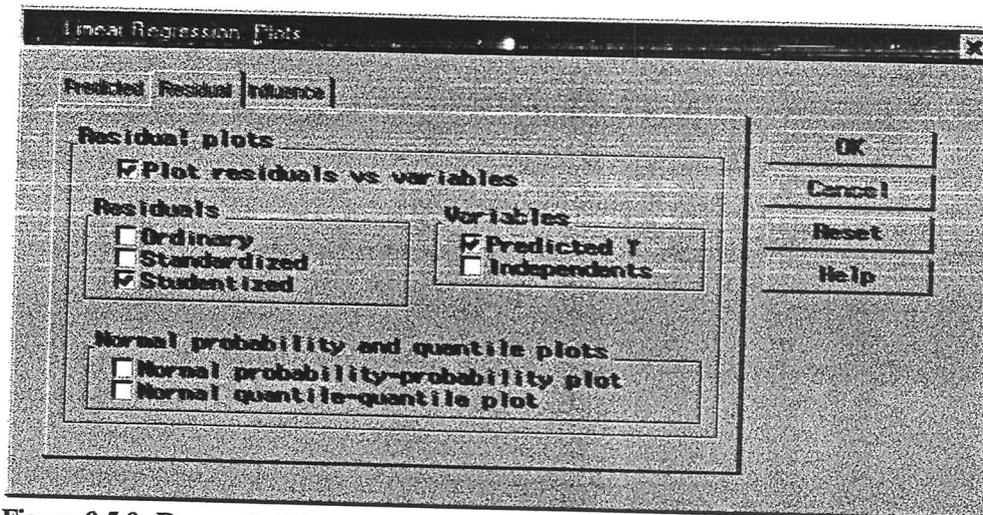


Figura 9.5.9: Regressão Linear: Diálogo Plots, Placa Residual

Um resíduo ordinário é a diferença entre a resposta observada e o valor predito para essa resposta. O resíduo padronizado é a razão do resíduo ordinário com o seu erro padrão. O resíduo estudentizado é o resíduo padronizado calculado com a observação corrente deletada da análise.

Clique OK no diálogo Linear Regression para executar a análise.

Examinando os Resultados

A figura 9.5.10 mostra a tabela de análise de variância e as estimativas dos parâmetros.

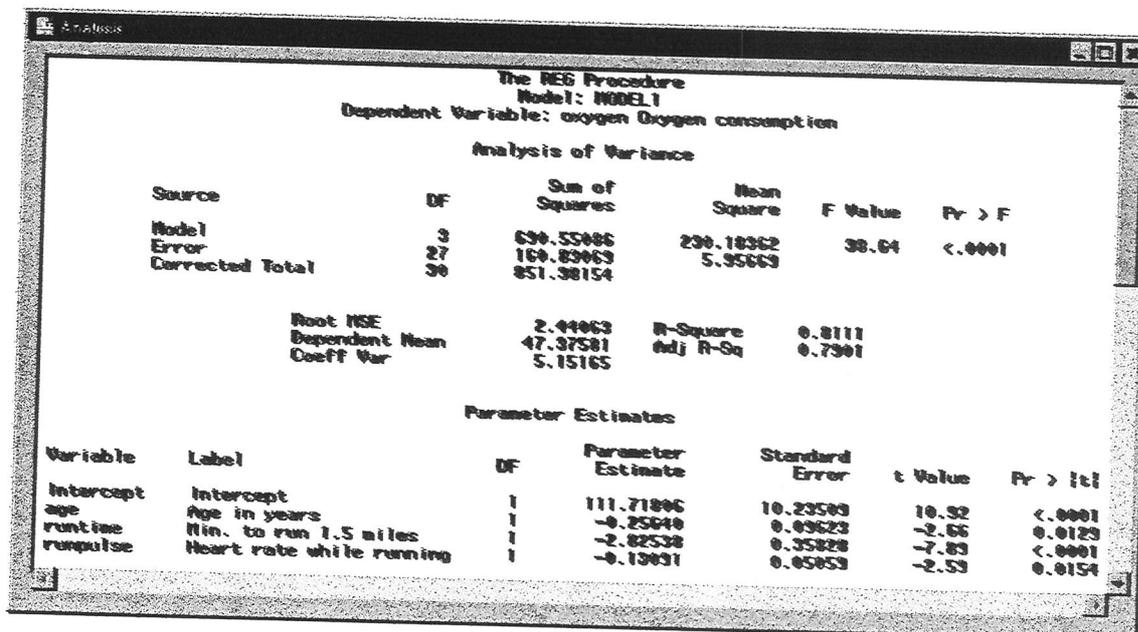


Figura 9.5.10: Regressão Linear: Tabela ANOVA e Estimativas dos Parâmetros.

Na tabela de análise de variância mostrada na figura 9.5.10, o valor F de 38.64 (com um p -value associado que é menor que 0.0001) indica o relacionamento significativo entre a variável dependente, oxigênio, e para no mínimo uma das variáveis explicativas. O valor R^2 indica que o modelo é responsável por 81% da variação no consumo de oxigênio.

A tabela "Parameter Estimates" lista os graus de liberdade, as estimativas dos parâmetros, e o erro padrão das estimativas. As duas colunas finais da tabela provêm dos valores t calculados e as probabilidades (p -values) associadas a obtenção de um valor absoluto t de tal magnitude. Cada p -value é menor que 0.05; deste modo, todas as estimativas dos parâmetros são significativas para o nível 5%. A equação ajustada para este modelo é a seguinte:

$$\text{oxygen} = 111.718 - 0.256 \times \text{age} - 2.825 \times \text{runtime} - 0.131 \times \text{runpulse}$$

A figura 9.5.11 mostra as estimativas por intervalo dos parâmetros e a tabela de diagnóstico de colinearidade.

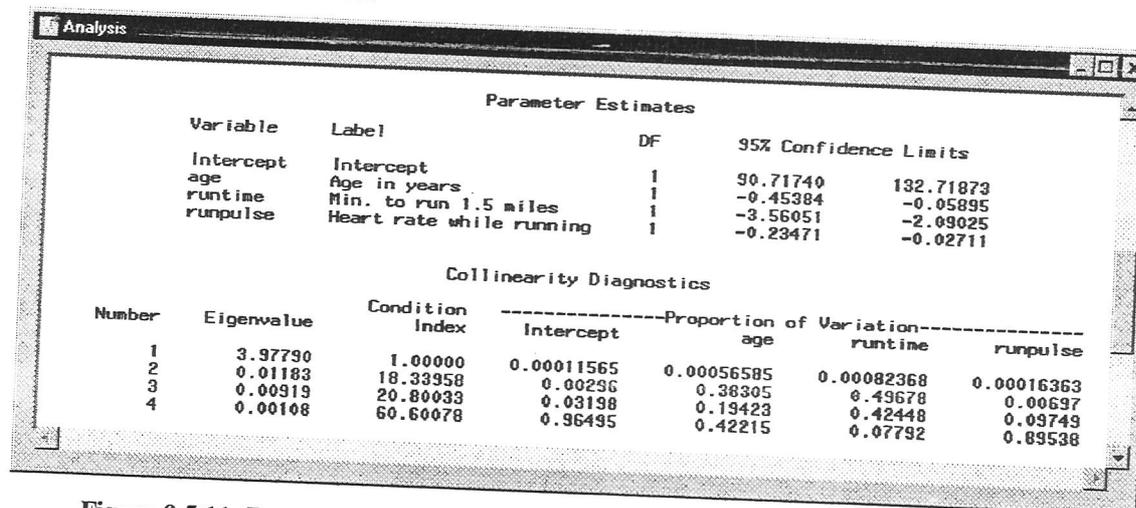


Figura 9.5.11: Regressão Linear: Limites de Confiança e Análise de Colinearidade

A tabela de diagnóstico de colinearidade mostra os eigenvalues (autovalores), o indicador de condição e a proporção de variação correspondente a cada estimativa. Geralmente, quando o indicador de condição é próximo à 10, há dependência frágil entre as estimativas de regressão. Quando o indicador é maior que 100, as estimativas podem ser instáveis, apresentando erros numéricos consideráveis. O diagnóstico mostrado na figura 9.5.11, apesar de indicarem dependências desfavoráveis entre as estimativas, não são tão excessivas a tal ponto de inviabilizar o modelo.

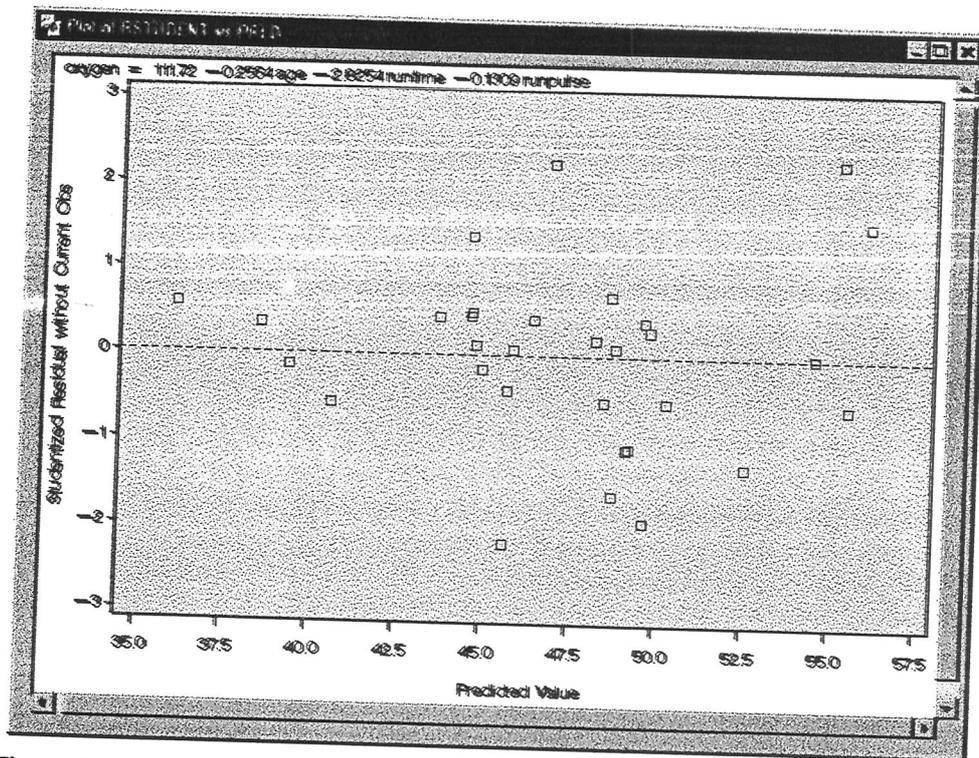


Figura 9.5.12: Regressão Linear: Gráfico de Resíduos Studentized versus Valores Preditos

O gráfico dos resíduos estudentizados versus valores preditos é mostrado na figura 9.5.12. Quando um modelo fornece um bom ajustamento e não viola nenhuma suposição do modelo, este tipo de gráfico de resíduos não exibe padrão ou tendência acentuados. A figura 9.5.12 não exibe nenhuma tendência, indicando um ajustamento adequado.

Exercício 23

Faça uma regressão linear com os dados do arquivo **calcario.xls**.

10. Solução dos Exercícios

Capítulo 8:

Gráficos

Exercício 8:

1. Abra o arquivo **chuva.sas**.
2. Clique em **Grahps** e em seguida escolha a opção **Bar Chart, vertical**.
3. Clique sobre a variável **Chuva** e a coloque no campo **Chart**.
4. No botão **Options** escolha o número de barras igual a 5
5. Na placa barras de valores clique sobre a variável **chuva** e em seguida no botão **Analisis**.
6. Clique em **OK**.
7. No botão **Titles** e escreva o título do gráfico: *Frequência de chuva*. Clique em **OK**.
8. Clique em **OK**.

Exercício 9:

1. Abra o arquivo **composicao.xls**.
2. Clique em **Grahps** e em seguida escolha a opção **Pie Chart**.
3. Clique sobre a variável **Compo** e a coloque no campo **Chart**.
4. No botão **Options** escolha o número de fatias igual a 5
5. Na placa barras de valores clique sobre a variável **PORCENTO** e em seguida no botão **Analisis**.
6. Clique em **OK**.
7. No botão **Titles** e escreva o título do gráfico: *Composição* . Clique em **OK**.
8. Clique em **OK**.

Exercício 10:

1. Abra o arquivo **mandioca.xls**.
2. Clique em **Grahps** e em seguida escolha a opção **Two Dimensional** do menu **Scatter Plot** .
3. Clique sobre a variável **MSR** e a coloque no campo **X Axis**. A variável **MST** deve ser posta no campo **Y Axis**. Se você quiser avaliar por variedade, coloque a variável **Variedade** no campo **Class**.
4. No botão **Display** você pode escolher a formatação do seu gráfico, quanto a cores e símbolos.
5. No botão **Titles** e escreva o título do gráfico: *Gráfico de Dispersão* . Clique em **OK**.
6. Clique em **OK** para executar a tarefa.

Capítulo 9:

9.1 Estatística Descritiva

Exercício 11:

1. Abra o banco de dados **pesovelo.xls**.
2. Selecione **Frequency Counts** do menu **Statistics**.
3. Posicione todas as variáveis na caixa **Frequencies**.
4. No botão **Plots** escolha a opção do barras verticais ou horizontais para exibir o gráfico de frequência de cada variável.
5. Clique em **OK** para executar a tarefa.

Exercício 12:

1. Abra o banco de dados **pesovelo.xls**.
2. Selecione **Summary Statistics** do menu **Statistics**.
3. Posicione as variáveis **PESOVELO** e **ALTURA** na caixa **Analisis**.
4. Se quiser realizar o cálculo separado para cada sexo, coloque a variável **SEXO** na caixa **Class**.
5. No botão **Statistics** escolha as estatísticas de seu interesse. Clique em **OK**.
6. No botão **Plots** escolha a opção de Histograma e/ou Box-Plot para a visualização do comportamento dos seus dados. Clique em **OK**.
7. Clique em **OK** para realizar a tarefa.

Exercício 13:

1. Vá até **Descriptive** no menu **Statistics** e clique em **Distributions**
2. Posicione a variável **Altura** na caixa **Analisis**.
3. No botão **Fit** você pode escolher algumas distribuições para serem ajustadas. Clique em **OK**.
4. No botão **Plots** escolha o tipo de gráfico do seu interesse. Clique em **OK**
5. Clique em **OK**.

Exercício 14:

1. Vá até **Descriptive** no menu **Statistics** e clique em **Correlations**.
2. Posicione as variáveis **Pesovelo** e **Altura** na caixa **Correlate**.
3. No botão **Plots** escolha o gráfico de dispersão com ou sem o intervalo de confiança. Clique em **OK**
4. No botão **Options** escolha algumas estatísticas que você queira calcular. Clique em **OK**.
5. Clique em **OK**.

9.2 Análise de Tabelas

Exemplo 15:

1. Abra o arquivo **broca.xls** salvo no seu diretório.
2. Abra o diálogo de Análise de Tabelas clicando em **Table Analysis** no menu **Statistics**.
3. Coloque a variável **Podverm** na caixa **Row** para esta se tornar uma variável de linha e a variável **Carvao** na caixa **Column** para esta se tornar uma variável de coluna.
4. No botão **Tables** você pode escolher quais valores serão apresentados na tabela. Selecione **observed** e **expected** para exibir os valores esperados e observados em cada célula.
5. No botão **Statistics**, escolha a estatística de seu interesse. Selecione **Chi-square statistics** para exibir o valor da estatística Qui-quadrado.
6. Clique em **OK** para construir a tabela.

9.3 Teste de Hipóteses

Exercício 16:

1. Abra o arquivo **pesovelo.xls**.
2. Clique em **Hypotesis Test** do menu **Statistics** e escolha a opção **One-Sample for a t-test for a mean**.
3. Posicione a variável **Pesovelo** na caixa **Variable** para efetuar o teste-t para esta variável.
4. No botão **Tests** selecione **Interval** para produzir um intervalo de 95% de confiança. Clique em **OK**.
5. Se quiser escolha um título para a análise clicando em **Titles**.
6. Clique em **OK**.
7. Repita o procedimento acima para a variável **Altura**.

Exercício 17.

1. Abra o arquivo **pantes_pdepois.xls**.
2. Clique em **Hypotesis Test** do menu **Statistics** e escolha a opção **Two-Sample Paired test for a mean**.
3. Posicione a variável **P_antes** na caixa **Group1** e **P_depois** na caixa **Group2**.
4. No botão **Tests** selecione **Interval** para produzir um intervalo de 95% de confiança. Clique em **OK**.
5. Se quiser escolha um título para a análise clicando em **Titles**.
6. Clique em **OK**.

Exercício 18.

1. Abra o arquivo **laranja.xls**.
2. Clique em **Hypotesis Test** do menu **Statistics** e escolha a opção **Two-Sample T-Test for a mean**.
3. Posicione a variável **Peso** na caixa **Dependent** e **Pomar** na caixa **Grup**.
4. No botão **Tests** selecione **Interval** para produzir um intervalo de 95% de confiança. Clique em **OK**.
5. Se quiser escolha um título para a análise clicando em **Titles**.
6. Clique em **OK**.

9.4 Análise de Variância

Exercício 19:

1. Abra o arquivo **pulgao.xls**.
2. Clique em **ANOVA** do menu **Statistics** e escolha a **One-Way ANOVA**.
3. Posicione a variável **NP** na caixa **Dependent** e **TRAT** na caixa **Independent**.
4. No botão **Tests** na placa **ANOVA** marque os testes de Bartlett, Brown-Forsythe e Levene na janela para os testes de igualdade de variâncias.
5. No botão **Means** na placa **comparisons** escolha o método **Tukey** no **Comparison method** e clique sobre a variável **TRAT** e clique em **ADD**(adicionar). Na placa **Breakdown** marque as estatísticas para variáveis quantitativas, por exemplo a média(**mean**) e o erro padrão(**std.error**). Clique em **OK**.
6. No botão **Plots** marque como o tipo de gráficos, aqueles de sua preferência, como por exemplo o **Box-plot**, o gráfico das médias e os gráficos de resíduos. Clique em **OK**.
7. Se quiser escolha um título para a análise clicando em **Titles**.
8. Clique em **OK**.

Exercício 20.

1. Abra o arquivo **anova_fatorial.xls**.
2. Clique em **ANOVA** do menu **Statistics** e escolha a **Factorial ANOVA**.
3. Posicione a variável **PROD** na caixa **Dependent** e **I** e **A** na caixa **Independent**.
4. No botão **Model** marque **I** e **A** na janela de independente e clique no botão **Cross** para incluir a interação **A*I** no modelo. Clique em **OK**.
5. No botão **Means**, na placa **LSMeans** marque **I**, **A** e **A*I** nos efeitos, assinale **compute p's for pairwise differences** e escolha o método **Tukey** como **adjustment method**. Clique em **OK**.
6. Clique em **Plots** na placa **Means**. Marque gráfico de medidas para efeitos principais e interação. Na placa **Residual** escolha sob a opção de plotar os resíduos vs variáveis. Em seguida, selecione Resíduos **ordinary** e variáveis **Predict Y** e **independents**.
7. Você cria e salva dados clicando no botão **Save Data**.
8. Se quiser escolha um título para a análise clicando em **Titles**.
9. Se você necessitar de uma análise separada para níveis de determinada variável você pode clicar no botão **Variable**.
10. Clique em **OK**.

Exercício 21.

1. Abra o arquivo **maca.xls**.
2. Clique em **ANOVA** do menu **Statistics** e escolha a **Linear Model**.
3. Posicione a variável **Peso** na caixa **Dependent** e **Trat** e **Bloco** na caixa **Class**.
4. No botão **Means** clique sobre a variável **Trat**, escolha o método **Tukey** e clique em **Add**. Clique em **OK**.
5. Clique em **Plots** e depois na placa **Residual** escolha sob a opção de plotar os resíduos vs variáveis. Em seguida, selecione Resíduos **ordinary** e variáveis **Predict Y** e **independents**.
6. Você cria e salva dados clicando no botão **Save Data**.
7. Se quiser escolha um título para a análise clicando em **Titles**.

8. Se você necessitar de uma análise separada para níveis de determinada variável você pode clicar no botão **Variable**.
9. Clique em **OK**

9.5 Regressão

Exercício 22.

1. Abra o arquivo **mandioca.xls**.
2. Clique em **Regression** do menu **Statistics** e escolha a **Simple**.
3. Posicione a variável **MSR** na caixa **Dependent** e **MST** na caixa **Explanatory**. Marque **Linear** no modelo.
4. Clique em **Statistics** e marque coeficientes de regressão padronizados e limites de confiança. Clique me **OK**.
5. Clique em **Plots**. Na placa **Predicted**, selecione a opção do gráfico de dispersão dos observados vs os preditos e dos observados vs variáveis independentes. Na placa **Residual** escolha sob a opção de plotar os resíduos vs variáveis. Em seguida, selecione **Resíduos ordinary** e variáveis **Predict Y**. Nos gráficos e quantis da probabilidade Normal selecione o gráfico de probabilidade Normal.
6. Você cria e salva dados clicando no botão **Save Data**.
7. Se quiser escolha um título para a análise clicando em **Titles**.
8. Você poderia proceder uma análise mais poderosa ou regressão segmentada para cada variável clicando em **Variable** e identificando variedade como variável de agrupamento.
9. Clique em **OK**

Exercício 23.

1. Abra o arquivo **calcario.xls**.
2. Clique em **Regression** do menu **Statistics** e escolha a **Linear**.
3. Posicione a variável **Rendimento** na caixa **Dependent** e **Calcario e fosforo** na caixa **Explanatory**.
4. Clique em **Statistics** e marque coeficientes de regressão padronizados e limites de confiança. Clique me **OK**.
5. Clique em **Plots**. Na placa **Predicted**, selecione a opção do gráfico de dispersão dos observados vs os preditos e dos observados vs variáveis independentes. Na placa **Residual** escolha sob a opção de plotar os resíduos vs variáveis. Em seguida, selecione **Resíduos ordinary** e variáveis **Predict Y**. Nos gráficos e quantis da probabilidade Normal selecione o gráfico de probabilidade Normal.
6. Você cria e salva dados clicando no botão **Save Data**.
7. Se quiser escolha um título para a análise clicando em **Titles**.
8. Clique em **OK**