

RELATÓRIO FINAL

“APLICAÇÕES DE TÉCNICAS ESTATÍSTICAS COM O USO DA INFORMÁTICA EM ORGANIZAÇÕES ”

Bolsista: Joseph Moutran Júnior

Orientador: Professor Moriz Blikstein

ÍNDICE

Introdução	03
Formulação do Problema da Pesquisa	06
Objetivos da Investigação	06
Hipóteses e conceitos-chave	06
Metodologia de investigação	07
Definição da Abrangência da Pesquisa	08
Administração em Segurança do Trabalho	08
Administração de Marketing	08
Trabalho Realizado	09
Metodologia aplicada na confecção das apostilas	11
Técnicas Estatísticas aplicadas em Segurança do Trabalho	12
Objetivo da Pesquisa na Área de Segurança do Trabalho	12
Principais Técnicas em Segurança do Trabalho	13
Qui-quadrado	13
Coeficiente de Correlação	14
Aplicações em Segurança do Trabalho	15
Qui-quadrado	15
Aplicação 01	15
Aplicação 02	17
Aplicação 03	18
Aplicação 04	20
Aplicação 05	21
Coeficiente de Correlação	22
Aplicação 01	22
Aplicação 02	24
Aplicação 03	26
Principais Técnicas em Administração de Marketing	27
Séries Temporais	27
Análise de Agrupamentos	28
Aplicações em Administração de Marketing	29
Previsão de Vendas	30
Análise de Agrupamentos	31
Conclusão	32
Relação das apostilas em anexo	35
Bibliografia	36

Introdução

A motivação para a realização deste trabalho surgiu durante minha experiência como monitor do LEPI - Laboratório de Ensino e Pesquisa de Informática - da EAESP, e como responsável pela preparação de material didático para o treinamento em Informática (4.000 horas) para funcionários do governo do Estado de São Paulo conforme contrato entre GVPEC e PRODESP, no período de dezembro de 94 a dezembro de 95; também como responsável pela área de treinamento em informática do banco General Motors. Para a preparação do material didático, consultamos manuais dos próprios fabricantes dos softwares bem como bibliografia dos principais autores. Constatamos que, de um modo geral, o material pesquisado é escrito por técnicos, em uma linguagem não acessível a todos os usuários comuns.

Daí, percebemos a necessidade de utilizarmos uma linguagem apoiada em conceitos de didática e em teorias de aprendizagem. As principais teorias utilizadas foram de Piaget (construtivismo) e Rogers (o professor como facilitador).

Piaget opina que o ensino tradicional na área da matemática e ciências afins tem efeitos prejudiciais para a aprendizagem das crianças. Uma vez desenvolvido este "trauma", a pessoa permanece com ele durante a vida adulta, o que viria a explicar a dificuldade das pessoas com uso da informática e da estatística. Isto acontece porque os métodos tradicionais se centram na transmissão direta do professor ao aluno e nas respostas corretas, ao invés de se centrar no pensamento autônomo e na construção de princípios matemáticos.

Para Piaget, novas ou diferentes formas de raciocínio não são exigidas. Não há uma atitude lógica-matemática especial. Há apenas a necessidade da aplicação das operações concretas e formais ao conteúdo da matéria.

A seguir temos os seis princípios do construtivismo DeVries e Kohlberg adaptados ao propósito desta pesquisa:

1. As estruturas psicológicas devem ser desenvolvidas antes que as questões numéricas sejam introduzidas. Se o aluno tenta raciocinar sobre questões numéricas antes de ter adquirido as estruturas lógicas-matemáticas relevantes aos conceitos matemáticos presentes nessas questões, os problemas não terão significado para ele. Daí a necessidade de se apresentar uma técnica estatística a partir de um problema real, problema este que o aluno lida em seu dia-a-dia. Isto irá interferir positivamente no processo de construção.

2. As estruturas psicológicas (esquemas) devem estar desenvolvidas antes que o simbolismo formal seja introduzido. O simbolismo nada mais é do que representações de conceitos. Os conceitos devem estar claros antes da introdução do simbolismo; caso contrário, ocorrerá uma confusão para o aluno. Mais uma vez se volta a necessidade de problemas reais do cotidiano do aluno. É preciso que ele entenda qual é o problema a ser resolvido e passo a passo como a técnica é capaz de resolvê-lo. Isto é válido tanto para a técnica estatística quanto para a técnica no uso do computador.

3. Não se deve enfatizar o conhecimento automatizado antes que a lógica implícita seja compreendida. A memorização antes da aquisição do conceito encoraja a memorização e não a construção e a compreensão. Este princípio é fortemente aplicado na construção das apostilas. Cada conceito é explicado em seu funcionamento e sua utilidade detalhadamente nas partes iniciais de cada apostila. À medida que o aluno avança na apostila, a repetição de cada conceito é feita sem tanto detalhes. Após várias repetições do mesmo conceito o aluno já é capaz de fazê-lo com um mínimo de explicações. Isto se dá pelo fato do aluno já ter internalizado e compreendido o conceito.

4. O aluno deve ter a oportunidade de inventar (construir) as relações lógicas-matemáticas em vez de simplesmente entrar em contato com o pensamento estruturado já pronto. A instrução bem sucedida é aquela que conduz à construção. Este princípio se reflete na utilização de problemas reais da área do profissional para a utilização de determinada técnica estatística. Assim, muitas vezes por analogia, o profissional terá a capacidade de desenvolver o mesmo conceito para outros problemas de seu ambiente de trabalho.

5. Os professores devem entender a natureza do erro dos alunos. Por definição, o desenvolvimento intelectual e matemático é cheio de “erros” e enganos. Os erros constituem uma parte inevitável do processo de construção em todas as áreas. O professor deve ter o erro como uma fonte de informação sobre o raciocínio do aluno. Princípio muito utilizado na confecção de nossas apostilas, pois sofrem constantes alterações de acordo com o feedback dos alunos.

6. Deve ser criada uma atmosfera própria para favorecer o ato de pensar. A forma típica de instrução consiste num esforço direto do professor em “despejar” nas crianças os fatos matemáticos e os procedimentos de cálculo. Normalmente os alunos são participantes passivos. Usamos intensamente a interatividade nas apostilas. Isto para tentar tirar o aluno de sua posição de passividade e melhorar a atmosfera que favoreça o ato de pensar. O uso intensivo de “telas” dos programas que o aluno está utilizando também contribui para esta atmosfera.

Uma outra corrente abordada nesta pesquisa é a corrente de Rogers. A teoria rogeriana é uma concepção positiva e otimista acerca da natureza humana. Rogers apóia sua orientação no princípio básico de que as pessoas possuem um potencial bastante rico para resolverem elas próprias suas dificuldades, desde que lhes seja proporcionada a oportunidade e a atmosfera adequadas. Neste contexto, Rogers vê o professor como um “facilitador”. Isto é, cabe ao aluno desenvolver-se por si só. O professor só irá facilitar este auto-desenvolvimento. Esta teoria tem relação com o sexto princípio mencionado acima da criação de uma atmosfera para favorecer o ato de pensar.

A partir destas teorias e idéias, estabelecemos as seguintes diretrizes na preparação do material didático, que é a principal contribuição desta pesquisa:

- 1) Usar diálogos simples, nos quais as informações necessárias aparecem em ordem natural e lógica. Nestes diálogos, não deve haver informações irrelevantes ou raramente utilizadas.
- 2) Usar a linguagem do usuário, com palavras, frases e conceitos que sejam familiares e de uso comum na área de domínio específico do usuário. Não usar termos técnicos específicos referentes ao software e à tecnologia da interface entre a máquina e o homem.
- 3) Ser consistente com os termos específicos utilizados, de maneira que seus significados não mudem de um capítulo para o outro.
- 4) Usar exemplos da área do usuário. De preferência, os exemplos devem ser relacionados com as tarefas realizadas pelo usuário.

Formulação do Problema da Pesquisa

As técnicas estatísticas são raramente utilizadas nas empresas por serem pouco conhecidas e de difícil operacionalização (cálculos geralmente muito trabalhosos).

Com o preço mais acessível de hardware e software, o uso da informática tornou-se muito disseminado nas organizações.

Nos softwares que possuem uma vasta gama de recursos para um tratamento estatístico de problemas, estes recursos são praticamente inaproveitados. Isto acaba gerando um hiato entre os recursos disponíveis para a resolução de problemas estatísticos e o uso efetivo dos mesmos.

Objetivos da investigação

Em uma primeira fase teve-se por objetivo pesquisar junto aos alunos do PEC as atuais aplicações estatísticas nas empresas.

Numa segunda fase objetivou-se a utilização tanto de softwares com funções estatísticas quanto softwares específicos de estatística na resolução das aplicações encontradas nas empresas pesquisadas.

Em uma terceira fase, teve-se por objetivo pesquisar novas aplicações estatísticas nas empresas.

Hipóteses e conceitos-chave

a) Geralmente, os cursos de estatística apresentam uma linguagem matemática e computacional muito pesada. Este fato gera uma desmotivação por parte do administrador na sua utilização.

b) Hoje, existe um hiato entre os recursos computacionais existentes na área de estatística e o seu uso efetivo.

c) Há uma necessidade de noções pedagógicas por parte dos especialistas em estatística.

Metodologia de investigação

Levantou-se , junto a alunos do PEC, situações problemas que permitiram tratamento estatístico.

A partir de tais problemas, foi elaborado um conjunto de exercícios-modelo, com as seguintes características pedagógicas:

- a) O texto foi escrito em um estilo não técnico;
- b) Os dados utilizados nos exercícios-modelo foram baseados em problemas reais, referentes a algumas áreas funcionais das empresas
- c) Os exercícios-modelo serviram de guia para mostrar como os resultados estatísticos podem ser usados no contexto da organização.

Definição da abrangência da pesquisa

O primeiro problema constatado na pesquisa foi o grande número de áreas com a possibilidade de aplicação de métodos estatísticos. Portanto, houve a necessidade de delimitação da abrangência da pesquisa. Foram escolhidas as seguintes áreas:

- Administração em Segurança do Trabalho
- Administração de Marketing

As justificativas para tanto foram:

Administração em Segurança do Trabalho:

Área que vem tendo um grande aumento de sua importância nos últimos anos devido a um aumento da conscientização do problema de segurança no trabalho. A escolha também deveu-se ao fato de o Orientador da pesquisa Prof. Moriz Blikstein estar desenvolvendo estudos nesta área e também por ministrar um curso sobre este assunto para alunos do PEC. O contato com estes alunos, profissionais da área de segurança do trabalho foi de extrema importância para a pesquisa, pois possibilitou uma vivência direta com as necessidades e os problemas da área.

Administração de Marketing

Uma área muito propícia para aplicações de estatística. Técnicas como Análise de Agrupamentos e regressão linear podem ser usadas com muita utilidade nesta área.

Trabalho realizado

No primeiro semestre de duração da bolsa foi dada prioridade a área de Segurança do Trabalho devido à possibilidade de contato com profissionais da área. A determinação das técnicas a serem usadas nesta área foi feita através de pesquisa com profissionais que estiveram no curso de Administração da Segurança do Trabalho. Professores deste curso também sugeriram algumas aplicações de Estatística em Segurança do Trabalho, como, por exemplo: Ergonomia e Epidemiologia Ocupacional

Das pesquisas realizadas, duas foram as técnicas que se mostraram mais úteis: Qui-quadrado e Correlação (regressão linear). Foram então criados e adaptados vários exemplos de aplicação destes dois métodos, baseados em problemas reais trazidos pelos profissionais da área de Segurança do Trabalho.

O próximo passo foi criar um material didático utilizando um exemplo simples de cada técnica. As apostilas foram escritas em uma linguagem acessível (não técnica) e este material foi apoiado no uso do software de planilha Excel, de fácil aprendizagem, custo acessível e bastante disseminado nas empresas. Estas apostilas seguem uma metodologia consagrada no ensino de cursos básicos de informática, baseado na experiência de quatro anos do bolsista na confecção deste tipo de material e na experiência do orientador na mesma área.

No caso do teste Qui-quadrado, foi criado um software para ser usado com a planilha Excel já que o mesmo não possui este recurso.

Além dos profissionais da área, foram fontes da pesquisa:

- Biblioteca da EAESP
- Bibliotecas da USP (Saúde Pública e FEA)
- Biblioteca da FUNDACENTRO
- Pesquisa através da Internet

Dentre estas fontes as que se revelaram mais frutíferas foram as pesquisas na FUNDACENTRO, órgão de renome na área de Segurança do Trabalho e o contato direto com os alunos do curso do PEC.

No segundo semestre, a pesquisa concentrou-se na área de Administração de Marketing. Foi feita uma pesquisa junto a profissionais da área, alunos do curso de Administração de Marketing do PEC, e junto ao Professor Jorge Motta, o que possibilitou a determinação das técnicas a serem utilizadas e a criação de exercícios-modelo que utilizassem estas técnicas.

Após isto, foram confeccionadas as apostilas utilizando a planilha eletrônica Excel.

Na utilização das técnicas em Administração de Marketing, também foi necessária a criação de software adicional para resolver os exercícios-modelo de séries temporais e também os de Análise de Agrupamentos.

Metodologia aplicada na confecção das apostilas:

A estatística é uma área tabu para muitos profissionais da área de administração. A informática é uma área que também causa traumas em muitas pessoas que necessitam da mesma como ferramenta de apoio em seu trabalho. Juntar os dois assuntos leva a uma área pouco explorada até na EAESP.

Portanto, criar material que ensine aos profissionais da administração em suas várias áreas a trabalhar com estatística usando o computador, deve ter como produto final apostilas extremamente claras e didáticas.

A experiência mostra que o tipo de apostila mais bem sucedido é o mais interativo possível. Apostilas e livros do tipo “Passo a passo” cada vez mais se consagram no mercado de ensino de informática. É justamente este tipo de apostila que se procurou desenvolver.

Neste método evita-se o tipo de didática abaixo:

Dentro do menu arquivo, escolha a opção Abrir e selecione o arquivo Teste.xls.

Em seu lugar fragmenta-se um único passo em vários passos menores para uma melhor compreensão do aluno:

- 1. Clique no menu Arquivo.*
- 2. Selecione a opção Abrir.*
- 3. Abra o arquivo Teste.xls clicando-o duas vezes.*

O uso de figuras também é fundamental neste tipo de metodologia. Proporciona ao aluno sempre uma localização segura de onde se encontra determinado comando ou botão. Facilita também o acompanhamento passo-a-passo do aluno muitas vezes não sendo necessária nem a presença de um instrutor.

No limite o uso de figuras leva a uma técnica de criação de apostilas apenas com figuras acompanhadas de um ou dois parágrafos.

Técnicas Estatísticas aplicadas em Segurança do Trabalho

Depois de pesquisa junto aos profissionais da área, foram identificados os seguintes problemas de inferência:

- Controle de coeficientes de incidência de acidentes e prevalência de ausências.
- Comparação de diferenças entre grupos
- Análise de séries históricas com tendência decrescente ou crescente
- Verificação da correlação entre medidas de segurança e índices de acidentes
- Previsão de resultados e cálculo do intervalo de confiança da previsão

Objetivo da pesquisa na Área de Segurança do Trabalho

Área de difícil penetração para profissionais de segurança do trabalho, a estatística tem sido o calcanhar de Aquiles de muitos estudos sobre acidentes.

Embora a tendência geral seja a de destacar o tratamento estatístico do planejamento e execução de pesquisas (confiando-o a um técnico especializado), o conhecimento de alguns princípios básicos de técnicas estatísticas é indispensável ao profissional da área pelos seguintes motivos:

1. Como uma etapa prévia do trabalho, enquanto não se tornar corrente, nas empresas brasileiras, a presença de um estatístico em cada “staff” de controle e avaliação.
2. Como subsídio técnico para o correto planejamento de pesquisas.

Principais técnicas em segurança do trabalho

Serão discutidas agora as principais aplicações na área de segurança do trabalho, dos principais testes de significância.

De acordo com a pesquisa, a grande maioria dos problemas de inferência em estatística de segurança do trabalho pode ser resolvida através dos seguintes testes básicos:

- Qui-quadrado
- Coeficiente de correlação (regressão linear)

Qui-quadrado

Por ser uma medida de dependência ou independência entre duas variáveis, suas aplicações na área de segurança do trabalho são inúmeras como, por exemplo, a análise de eficácia de uma campanha contra acidentes

Os softwares existentes só fornecem o valor final do qui-quadrado e não o valor por cada casela da tabela. Este valor é muito importante, pois permite localizar em que cruzamento da tabela ocorrem os maiores valores do qui-quadrado. E são estes valores que permitem localizar as correlações mais fortes entre as variáveis do estudo.

Pelo motivo acima, foi criado um programa no MS Excel para que se pudesse obter o valor do Qui-quadrado por cada casela da tabela. Tanto o programa quanto a apostila para sua utilização seguem em anexo. É importante ressaltar que o programa foi utilizado com pleno sucesso pelos alunos do curso de Administração em Segurança do Trabalho (PEC).

Coeficiente de Correlação

A teoria da Correlação estatística encontra larga aplicação na inferência de dados sobre administração em segurança do trabalho, pois permite a afirmação (ou negação) da mútua relação (de “causa-efeito” ou de “concomitância”) entre dois grupos de dados.

Exemplo: Índice de acidentes (y) e tempo de serviço (x).

Em geral, quanto maior o tempo de serviço (x) menor o índice de acidentes (y).

Esta correlação será diferente, de acordo com a empresa, com a categoria profissional estudada, com o tipo de acidente analisado, etc.

O conhecimento exato do grau dessa correlação seria útil, por exemplo, para avaliar a importância do treinamento na prevenção deste tipo de acidente, nas categorias profissionais estudadas.

É importante dispor de um instrumento matemático para medir parametricamente uma série de elementos indispensáveis para um bom conhecimento do fenômeno. Esses parâmetros são:

1. Equação de reta de tendência
2. Coeficiente de correlação entre y e x e grau de certeza dessa correlação
3. Intervalo de confiança em determinados pontos da reta de tendência

APLICAÇÕES EM SEGURANÇA DO TRABALHO

Nas aplicações propostas por profissionais da área de Segurança do Trabalho usando o teste de Qui-quadrado, observou-se que a maior parte destas aplicações tem como uma das variáveis envolvidas, o número de acidentes de trabalho ou a intensidade destes acidentes. Esta variável é muito importante para esta área, pois a relação desta com outras variáveis como nível de escolaridade, sexo, departamento, etc., é que vai determinar as políticas de prevenção dentro da empresa. Portanto, o exercício-modelo, utilizará esta variável.

QUI-QUADRADO

APLICAÇÃO 01

1 - Tema: Verificação da correlação entre depressão e sexo

A depressão que até poucas décadas era considerada uma manifestação puramente subjetiva hoje é aceita como sendo uma patologia provida de base biológica passível de tratamento medicamentoso específico.

A literatura mundial refere uma incidência maior de depressão no sexo feminino.

2 - Pesquisa de campo

Existe um questionário composto de doze questões divulgado pela Associação Americana de Psiquiatria que se propõe a identificar os indivíduos que podem ser considerados como portadores de depressão. A pessoa que responder afirmativamente quatro ou mais das doze questões é suspeita e deve consultar um profissional da área.

Para o presente trabalho foi aplicado um questionário a diferentes grupos de pessoas que têm em comum apenas o fato de serem moradoras da Grande São Paulo.

Com isto pretendeu-se inicialmente ter uma amostra geral na qual será verificada se há a incidência maior no sexo feminino ou não e que mais tarde possa servir de grupo de comparação ao se aplicar o mesmo questionário a amostras mais específicas seja por origem étnica, cultural ou por portar outra enfermidade quer congênita, quer adquirida.

O questionário foi aplicado a 91 pessoas, sendo 25 do sexo masculino e 66 do sexo feminino.

Considerou-se a possibilidade de depressão 'ausente' ou 'presente' de acordo com o número de respostas afirmativas respectivamente menor do que quatro ou igual e maior do que quatro.

Os resultados obtidos foram:

Pessoas do sexo feminino: presente 15, ausente 51

Pessoas do sexo masculino: presente 7, ausente 18

3. Formulação das hipóteses

- a. $H_0: X^2 = 0$ (hipótese nula) variáveis independentes
 $H_1: X^2 > 0$ há dependência entre sexo e depressão. Ela é maior no sexo feminino
- b. Região de Aceitação (R_a) de $H_0 : p = 3\%$
- c. Cálculo do X^2

	Masculino	Feminino	Total
presente	7	15	22
ausente	18	51	69
Total	25	66	91

Matriz do qui-quadrado:

	Masculino	Feminino
presente	0,15	0,06
ausente	0,05	0,02

O valor do qui-quadrado é:	0,27
O Número de Graus de Liberdade é:	1
O Valor de P é:	60,00%

Conclusão: As variáveis são independentes, pois o valor de P é alto, portanto H_1 foi rejeitada.

APLICAÇÃO 02

Objetivo:

Correlacionar os acidentes de trabalho 30 dias antes e 30 dias depois da Sipat (Semana Interna de Prevenção de Acidente do Trabalho) em três setores distintos da empresa, caracterizados por funcionários de alto nível intelectual (MD), médio nível (UM) e baixo nível (FM).

Fontes: Banco de Dados em Acidentes de Trabalho da Sismet.

H_0 : A Sipat não influencia no número de Acidentes.

H_1 : Dependência entre a Sipat e número de acidentes.

	FM	UM	MD	Total
30 dias antes Sipat	52	29	5	86
30 dias depois Sipat	40	23	13	76
Total	92	52	18	162

Matriz do qui-quadrado:

	FM	UM	MD
30 dias antes Sipat	0,20	0,07	2,17
30 dias depois Sipat	0,23	0,08	2,46

O valor do qui-quadrado é:	5,22
O Número de Graus de Liberdade é:	2
O Valor de P é:	7,37%

Conclusão: As variáveis são independentes, pois o valor de P é alto, portanto H_1 foi rejeitada.

Comentários: Verificando os resultados estatísticos percebe-se que as variáveis foram consideradas independentes, principalmente, por causa dos valores. A análise da estratégia usada para a Sipat mostra uma maneira bastante simples para se evitar acidentes, envolvendo pequenos jogos para as crianças e questionários para serem respondidos pelos funcionários, que podiam participar de sorteios de brindes oferecidos pela empresa. A criação de frases a serem colocadas na empresa e jogos de palavras cruzadas, também, faziam parte desta estratégia.

Acreditou-se que esta metodologia não se mostrou atraente na área de MD, onde o nível cultural é elevado. Houve, então, a necessidade da revisão estratégica da Sipat, oferecendo outras opções que se mostrassem atraentes a esta categoria de funcionários. Um exemplo desta revisão estratégica da Sipat foi a apresentação de diversos temas sobre prevenção de acidentes por grupos teatrais na empresa, atingindo acentuadamente os grupos intelectuais elevados.

APLICAÇÃO 03

1. Tema:

Verificar a correlação entre tempo de serviço e doenças de origem ocupacional.

Foi escolhido este tipo de doença por ser de carácter eminentemente ocupacional, isto é, decorrentes da tarefa que o trabalhador executa em sua atividade laboral.

2. Pesquisa de campo

Foi utilizado um questionário e tabuladas as seguintes variáveis:

- Funcionários com 0 a 5 anos de serviço e funcionários com 6 a 10 anos de serviço.
- Funcionários que relataram qualquer tipo de sintomatologia ocupacional e aqueles que não relataram qualquer sintomatologia.

3. Formulação das hipóteses.

H_0 = Variáveis independentes

H_1 = Variáveis dependentes

Anos	Sintomáticos	Assintomáticos	Total
0 a 5	41	35	76
6 a 10	15	18	33
Total	56	53	109

Matriz do qui-quadrado:

Anos	Sintomáticos	Assintomáticos
0 a 5	0,10	0,10
6 a 10	0,23	0,24

O valor do qui-quadrado é: 0,66

O Número de Graus de Liberdade é: 1

O Valor de P é: 41,50%

4. Conclusão

As variáveis são independentes, pois o valor de P é alto, portanto H_1 foi rejeitada.

5. Comentários

A hipótese pesquisada foi de que quanto mais tempo (em anos) um funcionário fica trabalhando, mais problemas de saúde ele poderá vir a ter.

Essa hipótese foi elaborada através de um questionário aplicado em uma fábrica de confecção, no momento do exame médico periódico, sendo que a maioria é do sexo feminino, onde se pesquisa a correlação e análise de sinais e sintomas de doenças ocupacionais versus a tarefa que esse trabalhador realiza.

Foram consideradas como sintomáticas as seguintes condições:

- dores nos ombros, braços, mãos e dedos;
- presença de cistos;
- dores na região cervical, lombar e sacra;
- queixas de dormência, formigamento, diminuição da força muscular e sensibilidade nas mãos.

Não só o fator tempo de serviço leva o trabalhador a ficar doente. Há outras variáveis que deveriam ser consideradas:

- Carga excessiva de trabalho, tarefas mais complexas;
- Pressão do tempo e tensão psicológica;
- Influência de problemas tensionais sérios que o trabalhador esteja vivendo no presente;
- Inexistência de condutas compatíveis com a segurança;

Para poder se estabelecer relação de dependência entre as variáveis expostas, necessitaria-se pesquisar outros itens tais como:

- Política da empresa diante da questão da saúde do trabalhador;
- Tipo de tarefa que o trabalhador realiza
- Biótipo de trabalhador para essa tarefa, antecedentes pessoais;
- Condições gerais de trabalho: postura, tensão, riscos ocupacionais.

APLICAÇÃO 05

Uma amostra da duração do afastamento por acidentes (dias) em relação aos horários da ocorrência.

Duração do afastamento Acidentes com duração > 15 dias
 Acidentes com duração <= 15 dias
 Acidentes sem afastamento com retorno no mesmo dia do acidente

Horário de Ocorrência 1h - 7h
 7h - 18h
 13h - 19h
 19h - 1h

Existem evidências de que a duração do afastamento e o horário da ocorrência estejam relacionados (dependência).

Hipóteses: $H_0 : X^2 = 0$ (Independentes)
 $H_1 : X^2 > 0$ (Dependentes)

	<= 15 dias	> 15 dias	Sem afastam.	Total
01 - 07 hs.	2	1	27	30
07 - 13 hs.	10	8	46	64
13 - 19 hs.	7	11	69	87
19 - 01 hs.	2	2	37	41
Total	21	22	179	222

Matriz do qui-quadrado:

	<= 15 dias	> 15 dias	Sem afastam.
01 - 07 hs.	0,25	1,31	0,33
07 - 13 hs.	2,57	0,43	0,61
13 - 19 hs.	0,18	0,66	0,02
19 - 01 hs.	0,91	1,05	0,47

O valor do qui-quadrado é: 8,78

O Número de Graus de Liberdade é: 6

O Valor de P é: 18,62%

Conclusão: As variáveis são independentes, pois o valor de P é alto, portanto H_1 foi rejeitada.

REGRESSÃO LINEAR (coeficiente de correlação)**APLICAÇÃO 01**

Justificativa - A empresa onde foi realizada a pesquisa é um banco onde o número de acidentes de trabalho devido às Lesões por Esforços Repetitivos - LER - é progressivo.

Em observação aos registros de Comunicações de Acidentes de Trabalho - CAT -, notou-se um número acentuado de pessoas do sexo feminino acometidas por LER. Outro dado que ficou destacado é que a maioria dos casos, tanto do sexo feminino quanto do sexo masculino, são empregados detentores de função executiva ou gerencial.

Dessa forma a preocupação foi de verificar se estes dados estavam relacionados, ou seja, procurar saber se o fator genético no que diz respeito ao sexo é desencadeante de alguma afecção por LER, e se o exercício da função influi neste dado.

Para tanto foi investigado os números de acidentes de trabalho por LER de algumas agências, principalmente aquelas com número de casos elevados. Também foram procurados os empregados com função e o tempo de exercício desta função, uma vez que acreditou-se que este fator também estaria interligado.

Assim, as variáveis são as seguintes:

Y = Número de Acidentes de trabalho por LER

X1 = Sexo Feminino

X2 = Empregados com função

X3 = Tempo de Exercício da Função

AGÊNCIAS	NR.AT (LER)	FEM	C/FUNÇÃO	TEMPO FUN
A	7	4	5	1,8
B	10	9	6	6
C	9	8	5	2,2
D	6	6	6	4,3
E	7	6	5	6,2
F	9	9	8	4,4
G	5	5	3	1,6
H	6	3	4	2
I	11	9	5	3,6
J	7	5	5	1,6
L	8	5	6	3,3

REGRESSÃO SIMPLES

RESUMO DOS RESULTADOS				
<i>Estatística de regressão</i>		<i>Coefficientes</i>		<i>valor-P</i>
R múltiplo	0,82600262	Interseção	3,26968504	0,01345603
R-Quadrado	0,68228032	Variável X 1	0,71062992	0,00172982
R-quadrado ajust.	0,64697814			
Erro padrão	1,09849575			
Observações	11			

REGRESSÃO MÚLTIPLA

RESUMO DOS RESULTADOS				
<i>Estatística de regressão</i>		<i>Coefficientes</i>		<i>valor-P</i>
R múltiplo	0,832958	Interseção	2,87382669	0,13175
R-Quadrado	0,69381903	Variável X 1	0,71848758	0,01756
R-quadrado ajust.	0,56259861	Variável X 2	0,1482126	0,7127
Erro padrão	1,22274996	Variável X 3	-0,12929913	0,6678
Observações	11			

Os resultados das regressões (R^2 ajustado = 56%) demonstram que não existe correlação forte entre as variáveis. No caso da ocorrência do número de acidentes de trabalho por LER ser maior em mulheres, podem haver duas razões: uma delas é que existem mais mulheres na empresa e a outra é que elas são as que mais exercem serviços que exigem esforços repetitivos.

Possíveis desdobramentos

A área de saúde e segurança da empresa iniciou um Programa de Ações preventivas das LER, que inclui a realização de palestras nas agências, distribuição de folhetos informativos, implantação dos 10 minutos de pausa a cada 50 minutos trabalhados (conforme a NR17) entre outras.

Assim pode-se acompanhar a implantação deste Programa nas agências escolhidas para o exercício da regressão e saber se tais ações são uma medida real de diminuição do número de Acidentes de Trabalho por LER.

Para pesquisar as ações curativas da LER, pode-se incluir dados relativos aos diversos tratamentos realizados pelos empregados doentes das agências escolhidas. Desta forma, poderemos diferenciar os tratamentos e os resultados obtidos, separando as pessoas que fazem acupuntura, fisioterapia, laserterapia, entre outros tratamentos.

Outro possível desdobramento se refere à questão ambiental: visitar as agências escolhidas para a regressão e pesquisar em cada uma delas dados de iluminação, ventilação, existência de poeira, poluição sonora ou visual. Com esta variável podemos estabelecer se existe relação entre o número de acidentes de trabalho por LER e má qualidade do ambiente.

APLICAÇÃO 02

Tema: Relação existente entre número de horas extras e número de acidentes no trabalho

Escolha das variáveis:

Y - Número de Acidentes

X1 - Número de Horas extras

X2 - Número de Funcionários

As variáveis foram escolhidas para verificar se há correlação entre aumento de número de acidentes de trabalho com aumento de horas extras, e a partir dos resultados, se necessário implantar ações preventivas.

Mês	Nr. de Acidentes	Nr de horas Extras	Nr. de Funcionários
Jan	1	19364	890
Fev	5	15444	890
Mar	6	7755	899
Abr	7	8471	911
Maio	7	4013	911
Jun	0	3754	912
Jul	2	4518	931
Ago	1	5880	915
Set	3	6550	925

REGRESSÃO LINEAR SIMPLES

RESUMO DOS RESULTADOS

<i>Estatística de regressão</i>		<i>Coeficientes</i>		<i>valor-P</i>
R múltiplo	0,03575446	Interseção	3,70759662	0,08921913
R-Quadrado	0,00127838	Variável X	-1,8065E-05	0,92723916
R-quadrado ajust.	0,14139614	1		
Erro padrão	2,93124145			
Observações	9			

REGRESSÃO LINEAR MÚLTIPLA**RESUMO DOS RESULTADOS**

<i>Estatística de regressão</i>		<i>Coeficientes</i>		<i>valor-P</i>
R múltiplo	0,35604331	Interseção	107,481122	0,37338426
R-Quadrado	0,12676684	Variável X 1	-0,0002499	0,45836549
R-quadrado ajust.	0,16431088	Variável X 2	-0,11197468	0,38894344
Erro padrão	2,96051914			
Observações	9			

Análise do Resultado da Regressão Simples

O R^2 da regressão simples mostrou a não existência de relação entre as variáveis escolhidas. Para concluir-se que havia correlação seria necessário que o valor de R^2 fosse maior que 80%.

Análise do Resultado da Regressão Múltipla

Embora o R^2 tenha aumentado em relação à regressão simples, conclui-se que as variáveis não têm relação entre si.

Possíveis desdobramentos:

Analisar um número maior de períodos, pois o período analisado não permite julgar se o modelo é bom ou ruim mesmo com estes resultados de R^2 e p.

Fazer esta análise em outras empresas, com graus de riscos diferentes e que fabriquem outros tipos de produtos.

APLICAÇÃO 03

Tema: Relação entre as variáveis Acidentes de Trabalho e Funcionários com EPI.

Foi realizado um estudo durante o período de 03 anos com 30 funcionários de vários setores de uma empresa e foi constatado o seguinte:

DOCAS	ACID. TRAB.	FUNCION. C/ EPI
A	7	20
B	1	25
C	5	27
D	4	28
E	10	15
F	8	12

RESUMO DOS RESULTADOS				
<i>Estatística de regressão</i>		<i>Coefficientes valor-P</i>		
R múltiplo	0,78536964	Interseção	13,8453922	0,01351452
R-Quadrado	0,61680546	Variável X 1	-0,37852247	0,06415569
R-quadrado ajust.	0,52100683			
Erro padrão	2,20675415			
Observações	6			

O valor de R^2 ajustado = 0,52 é significativo, entretanto não foi forte o suficiente para afirmar-se que há uma boa relação entre as duas variáveis.

PRINCIPAIS TÉCNICAS EM ADMINISTRAÇÃO DE MARKETING

Junto a profissionais da área de Marketing definiu-se duas técnicas de grande importância prática: Séries Temporais e Análise de Agrupamentos.

Séries Temporais:

Uma série temporal é um conjunto de medidas , ordenadas no tempo. Ela possui características especiais que necessitam de novos métodos estatísticos para sua análise. Diferente das técnicas estatísticas baseadas em amostragem aleatória, aqui a observação de um período estará relacionada com a observação dos períodos adjacentes. Outra característica das séries temporais é a sazonalidade. Por exemplo se no ano de 1996, dezembro foi um bom mês para a venda de cerveja, é grande a probabilidade de que dezembro de 1997 seja também um bom mês para a venda do produto.

Um modo de analisar o comportamento de uma série temporal é analisá-la em seus componentes. Quatro podem ser os componentes de uma série temporal. Tendência, sazonalidade, comportamento cíclico e comportamento irregular. A tendência de uma série indica sua capacidade de aumentar ou diminuir regularmente por um certo período de tempo. Um exemplo seria o número de filhos de casais brasileiros que vem diminuindo paulatinamente ao longo do tempo. O comportamento da sazonalidade consiste de padrões repetidos em intervalos de tempo determinados. Por exemplo, o comportamento da venda de cerveja é uma série com sazonalidade. Ano após ano esta série tem como característica o aumento nos meses de verão e sua queda nos meses de inverno. Em adição muitas dos negócios na vida real apresentam padrões oscilatórios ou cíclicos que não têm conexão com o comportamento sazonal. Este padrões podem não serem regulares. Por último ainda se tem a irregularidade da série que é devida a um grande número de fatores. Estes fatores influenciam a série e são imprevisíveis.

Foram desenvolvidos a partir da planilhas Excel programas para os modelos de Médias Móveis, Holt e Holt-Winters.

O modelo de Holt é capaz de analisar séries com o componente da tendência.

Os modelos de Médias Móveis e Holt-Winters são capazes de analisar séries com tendência e sazonalidade.

Análise de Agrupamentos:

Análise de Agrupamentos é o nome dado a um conjunto de técnicas cujo objetivo é agrupar os indivíduos de um conjunto em subgrupos homogêneos. A similaridade entre indivíduos de um mesmo grupo será maior que a similaridade entre indivíduos de agrupamentos distintos.

A classificação é feita a partir de características que são definidas para cada indivíduo com objetivo de através delas diferenciá-los em conjuntos homogêneos. Por exemplo, podemos agrupar alunos de uma classe de acordo com características como Notas, frequência, idade, etc.

Uma das aplicações da Análise de Agrupamentos que mais interessará a esta pesquisa é a segmentação de mercados. Os consumidores com suas várias características (variáveis) serão classificados em agrupamentos (segmentos de mercado). A análise do perfil dos consumidores de cada segmento, baseada em características demográficas e comportamentais, permitirá planejar as estratégias de mercado para os diferentes segmentos.

Outras aplicações são: identificação de mercados potenciais para determinados produtos, determinação de mercados idênticos em países diferentes e obtenção de grupos de consumidores que possam servir de referência na previsão de vendas.

APLICAÇÕES EM ADMINISTRAÇÃO DE MARKETING

As aplicações das técnicas previsão na Administração de Marketing são inúmeras conforme pesquisa feita junta ao Professor Jorge Motta. Todas aplicações abaixo, originaram-se de problemas reais trazidos por alunos do Professor Motta.

1. Previsão de vendas para a companhia como um todo, por linha de produtos, por canal de distribuição e por classe de clientes.
2. Estimação do lucro de um exercício fiscal.
3. Estimação da participação de mercado de um produto ou serviço.
4. Estimação das quotas de vendas por região geográfica e por vendedor.
5. Previsão dos orçamentos de gastos por tipo de esforço mercadológico (propaganda, promoção de vendas, venda pessoal), por linha de produtos e por região geográfica.
6. Previsão de custos e despesas de uma determinada linha de produtos ou serviços.
7. Estimação do potencial de mercado ou de vendas.
8. Previsão dos preços futuros de determinados produtos ou serviços.
9. Determinação da relação entre as características sócio-econômicas de uma família e quantidade de um dado produto que ela tem capacidade de adquirir no período de um ano.
10. Determinação da extensão da proporção de compras de produtos ou serviços identificados por marcas privadas por uma família associada às características psicológicas do marido ou da esposa.
11. Verificar, se, em termos médios, uma redução de 10 por cento no preço de venda de uma determinada marca de produto ou serviço terá um impacto maior na sua participação de mercado do que um aumento da verba de propaganda ao varejo em 10 por cento.
12. Determinação da utilidade de medidas econômicas gerais como produto nacional bruto, porcentagem de desemprego, etc., como base para previsão das vendas de uma determinada indústria ou setor econômico.

PREVISÃO DE VENDAS**APLICAÇÃO 01**

Iremos usar como modelo o seguinte exercício de autoria do Prof. Jorge Motta.:

A empresa Malhasmoda Indústria e Comércio de Malhas Ltda. fabrica e comercializa malhas para adultos e crianças desde sua fundação em 1957. Em dezembro de 1996, sua diretoria comercial estava empenhada em estimar as vendas de janeiro de 1997. Os dados apresentados a seguir abrangem o período compreendido entre janeiro de 1993 e dezembro de 1996. Os valores são expressos em US\$ 1.000 dólares. Os dados seguem abaixo:

	ANO			
<u>MÊS</u>	<u>1993</u>	<u>1994</u>	<u>1995</u>	<u>1996</u>
janeiro	4674	3811	5419	5157
fevereiro	5364	4081	6106	6408
março	5589	4772	6401	6936
abril	6315	5361	7343	7581
maio	6698	6650	7515	8536
junho	7098	7096	7697	8499
julho	5971	5793	6846	7004
agosto	6212	6430	7308	7958
setembro	6436	6860	7620	8217
outubro	5550	7082	6479	7682
novembro	5502	6812	4413	7124
dezembro	<u>3592</u>	<u>4589</u>	<u>3467</u>	<u>4771</u>
Total	69001	69337	76614	85873

Um exame atento do quadro apresentado acima revela que as vendas da Malhasmoda Indústria e Comércio de Malhas Ltda. tendem a aumentar no período de quatro anos, porém variam enormemente dentro de cada ano. A direção geral das vendas de um mês para o seguinte relaciona-se com variações ocorridas durante o mesmo período em outros anos. Por conseguinte, cada mês dentro do ano parece conter influência sazonal. As vendas são registradas em preços constantes, uma vez que alterações de preço influenciaram inevitavelmente a análise.

(Dados colhidos pelo Professor Jorge Motta).

ANÁLISE DE AGRUPAMENTOS

APLICAÇÃO 01

A companhia Hal é uma das grandes empresas em sua área e desenvolve equipamentos tecnologicamente avançados para uma grande variedade de aplicações. Uma de suas divisões é a de equipamentos de escritório. Tradicionalmente seus dois maiores competidores neste ramo foram as empresas Kraft e Calt. Entretanto nos últimos dez anos estas duas empresas vem ganhando mercado em relação à Hal.

Com o objetivo de manter esta divisão da empresa lucrativa, seus executivos delinearam quatro estratégias para promover as vendas: (1) Identificar as principais contas da empresa. (2) Examinar o processo de decisão de compra destas empresas. (3) determinar os serviços de vendas críticos procurados por estas empresas. (4) medir a performance da Hal e de suas competidoras nestes serviços.

Para implementar estas quatro estratégias, foi desenvolvido um questionário onde foram colocadas várias questões para que o comprador identificasse as mais importantes e as menos importantes. Também se encontrava neste questionário, uma avaliação de qual a performance de cada uma das três empresas em cada questão.

A partir da volta dos questionários e da tabulação dos mesmos. Foram selecionadas as características mais importantes de acordo com os dados obtidos nos questionários. Estas características foram usadas como variáveis para proceder-se a Análise de Agrupamentos. Com estas variáveis, utilizou-se um software adequado, como o Systat, e as empresas foram classificadas de acordo com suas respostas. Os clusters de da Hal e das outras duas empresas foram identificados e esta análise de clusters permitiu a implantação das demais estratégias.

CONCLUSÃO:

Após duas aulas nas quais foram ministradas as técnicas de regressão linear e séries temporais (modelo de Holt-Winters), foi feita uma enquete com os alunos, com base na seguinte pergunta:

“Tendo recursos estatísticos numa planilha de fácil utilização apoiada em material didático em linguagem intuitiva (não técnica), você se sente motivado a utilizar tais recursos?”

100% dos alunos entrevistados, responderam sim à pergunta. Alguns comentários feitos a tal questão seguem abaixo:

Creio que ter ferramentas que norteiem a área de vendas e marketing em suas simulações é muito bom.

Desde que de fácil entendimento e orientado para resultados bem próximos da realidade.

Desde que conheça a lógica contida

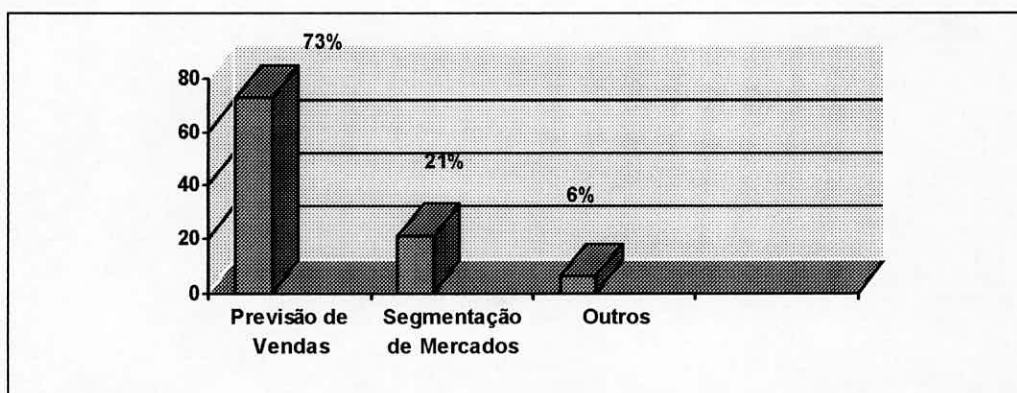
Estes recursos seriam um dos apoios para uma tomada de decisão.

Auxilia na definição das estratégias e planejamento

A dificuldade reside em não conhecer as técnicas estatísticas.

Eu me sinto motivado a utilizar tais recursos.

Na segunda questão, podemos ver quais são as principais áreas de interesse dos alunos:



- Previsão de vendas foi o assunto de maior interesse dos entrevistados. 73% gostaria de utilizar as técnicas aprendidas neste problema.
- 11% dos entrevistados mostraram interesse na área de Segmentação de Mercados.

- Previsão de Demanda
- Estimativas de Custos
- PCP
- Custo gráfico
- Tabulação de Pesquisas
- Gráficos
- Margem de Lucro
- % de Vendas
- Planilha de Controle e cálculo da lucratividade dos produtos
- Previsão de Produção
- Maximização dos recursos disponíveis
- Evolução de um cadastro de clientes.
- Contabilidade
- Despesas X Receitas
- Previsão de Vendas tendo como base o custo de uma commodity
- Planejamento micro e macroeconômico
- Custos - Vendas
- Avaliação da lucratividade X Boa Remuneração para Vendas
- Determinação do Ponto de equilíbrio
- Dimensionamento da força de Vendas
- Simulação de Vendas Globais de Produtos

Os resultados desta enquete vieram a corroborar as nossas premissas iniciais. O primeiro princípio de De Vries e Kohlberg - as estruturas psicológicas devem ser desenvolvidas antes que as questões numéricas - resultou na proposição de problemas concretos do dia-a-dia do profissional em administração. Isto levou o aluno a se sentir motivado, pois as técnicas por ele aprendidas fazem sentido. Mais até do que isto são percebidas como ferramentas necessárias no seu trabalho.

O próprio simbolismo formal foi sendo introduzido aos poucos e de preferência após as estruturas psicológicas estarem desenvolvidas. Isto é, após o aluno entender realmente a utilidade da técnica. A utilização de uma linguagem não-técnica (intuitiva) permitiu que os alunos pudessem chegar à resolução de um problema de uma maneira mais rápida. Esta rapidez e facilidade no uso das ferramentas inclusive gerou comentários de que as técnicas apresentadas eram simples. Antes do uso deste material, conforme o Professor Motta, os alunos achavam as técnicas apresentadas complicadas.

De acordo com o terceiro princípio de De Vries e Kohlberg foi enfatizada a compreensão lógica do problema em detrimento da memorização. Esta só foi utilizada nos passos finais de cada apostila. Neste ponto o aluno já tinha entendido a lógica do problema e era capaz de automatizar os algoritmos para a resolução do mesmo.

Outro princípio aplicado e que foi comprovado pela enquete, foi a necessidade de se criar uma atmosfera própria para favorecer o ato de pensar. Podemos comprovar a sua correta aplicação através dos vários novos exemplos de aplicação sugeridas pelos alunos.

Relação das apostilas em anexo.

Foi confeccionada uma apostila sobre cada um dos assuntos acima relacionados (desvio-padrão, qui-quadrado e regressão linear, séries temporais, análise de agrupamentos).

Entre os vários softwares disponíveis para o uso estatístico dois foram escolhidos: o Microsoft Excel e o Systat. O primeiro pela facilidade de uso e o fato de ser largamente utilizado em todo o mundo. O segundo por ser um software especialista na área de estatística e que se encontra a disposição nos laboratórios da EAESP.

Entretanto, após a confecção de algumas apostilas para o Systat, chegou-se a conclusão de que este software não possui as facilidades de uso do Microsoft Excel e decidiu-se pelo abandono da confecção de apostilas sobre o mesmo.

Em anexo se encontram as seguintes apostilas:

- Qui-quadrado utilizando o programa desenvolvido pelo bolsista em anexo para a versão 5.0 ou superior do MS Excel.
- Regressão linear utilizando o MS Excel
- Análise de agrupamentos utilizando o Systat
- Séries temporais: Modelo de Holt-Winters

7. Bibliografia

Berrigan, J., e Carl F., 1994, Marketing de Segmentação, Makron Books

Buschinelli, Jose T. P., Lyz E. R., Raquel M. R.(orgs), 1993, Isto é trabalho de gente?, Ed. Vozes

Bussab, Wilton O. e Pedro A. M., 1987, Estatística Básica, Atual Editora

Churchill Jr., G, 1995, Marketing Research - Methodological Foundations, The Dryden Press

F. Grundy M. D. e Reinke, W. A., 1974, Procedimentos em pesquisa de saúde, Federação Brasileira de Hospitais

Gentner, D, and Albert L. S (editores)., 1983, Mental Models, Lawrence Erlbaum Associates Publishers

Gordon, A., Michael Hacker, and Marc de V., 1993, Advanced Educational Technology in Technology Education, Springer - Verlag

Levine, David M., Patricia P.R., and Mark L.B., 1995, Business Statistics for Quality and Productivity, Prentice - Hall International Inc.

Lapponi, Juan Carlos, 1997, Estatística usando Excel, Lapponi Editora

Ragsdale, C. T., 1995, Spreadsheet Modeling and Decision Analysis, Course Technology

Van der Veer, G, and Mulder, G (editores), 1988, Human - Computer Interaction: Psychonomic Aspects, Springer - Verlag

Wilkinson, L., 1992, Manual do Systat, Systat Incorporation

QUI-QUADRADO USANDO O EXCEL

Autor: Joseph Moutran Júnior
Orientador: Prof. Moriz Blikstein

Autor: Joseph Moutran Júnior
Orientador: Prof. Moriz Blikstein

APOSTILA DE QUI-QUADRADO USANDO O EXCEL

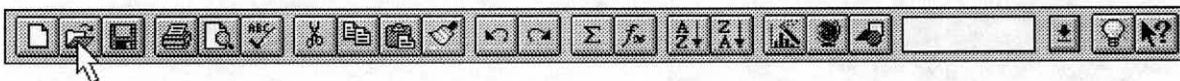
Vamos abrir o programa Excel. Para isso dê dois clique no ícone mostrado abaixo:



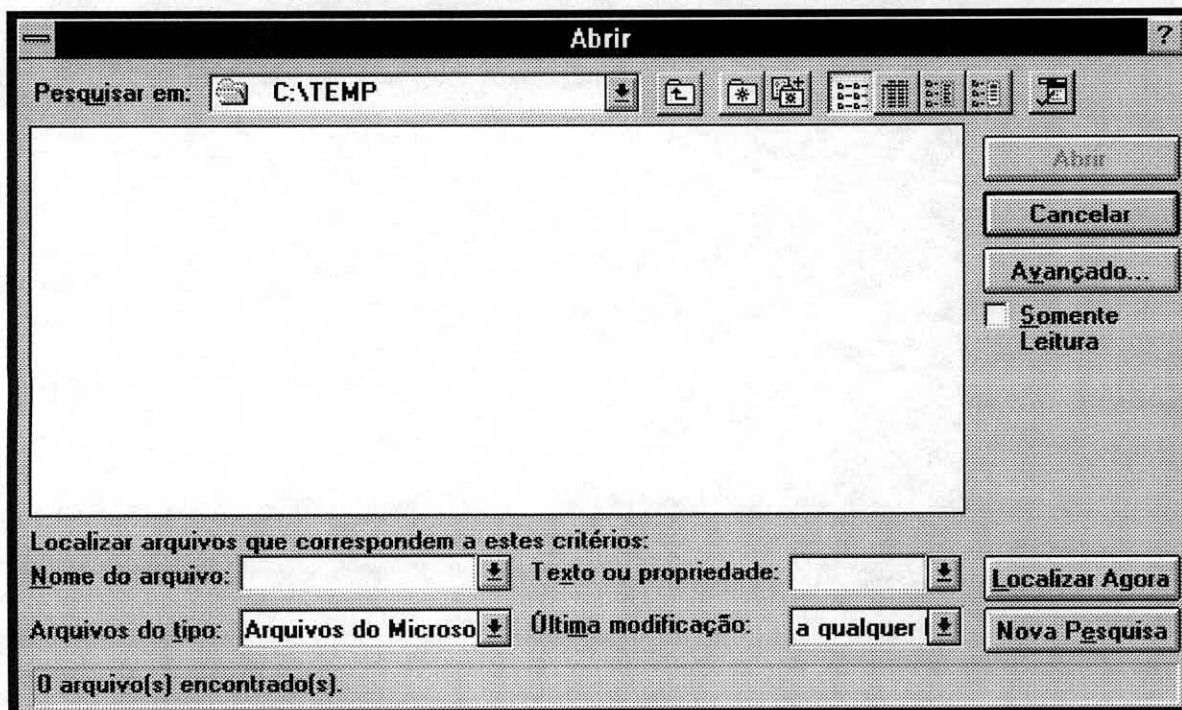
Na janela que se abre, dê um duplo clique no ícone do Microsoft Excel mostrado abaixo:



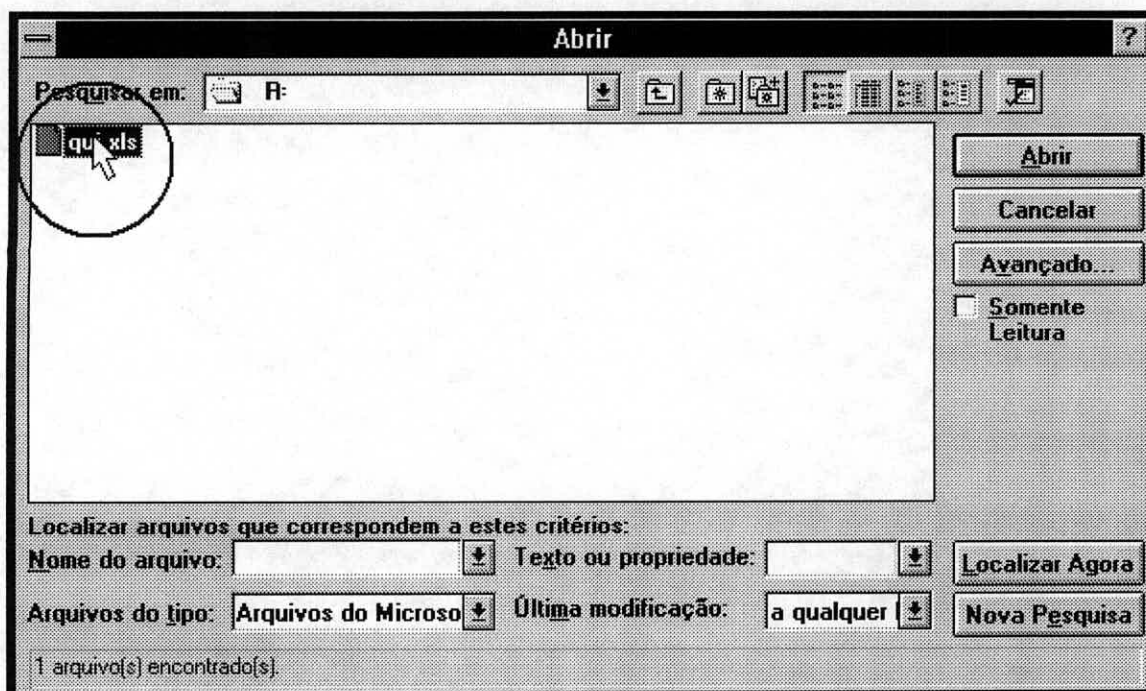
Em primeiro lugar, abra o arquivo **QULXLS** que se encontra na unidade **P**. Para isso, clique com o mouse no botão **Abrir**:



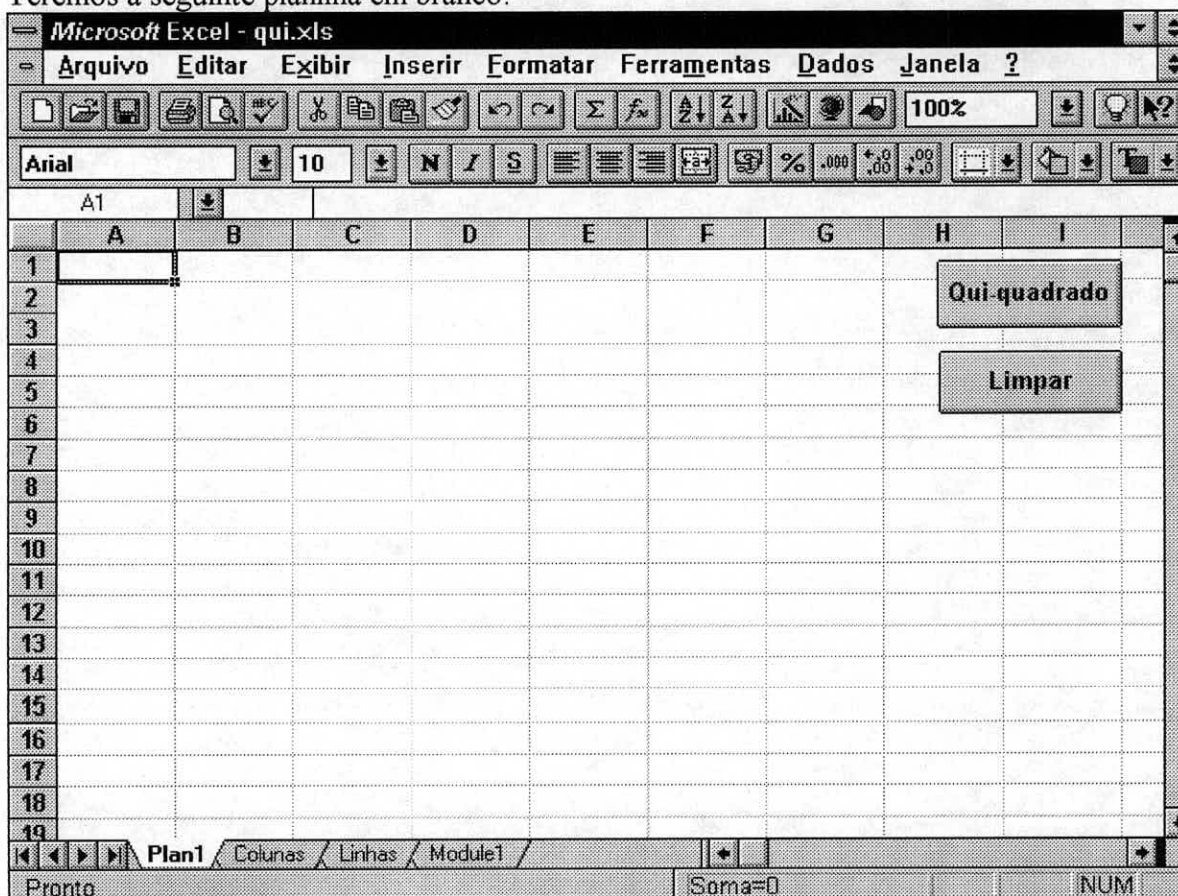
Aparece a seguinte Caixa de diálogo:



Selecione a **Unidade A:**. Aparecerá o arquivo **qui.xls**:



Teremos a seguinte planilha em branco:



Deveremos digitar os dados com as seguintes restrições:

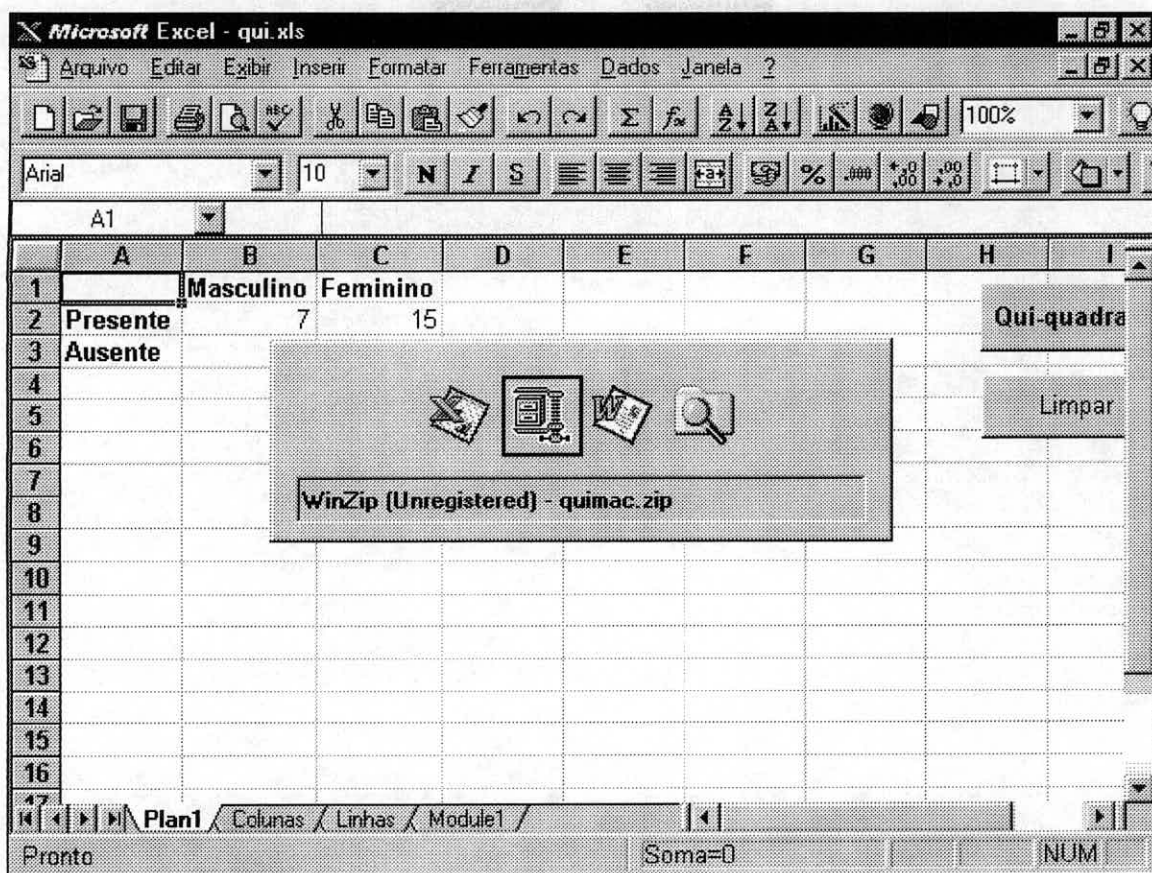
- Na primeira linha, e somente na primeira linha, deverão ser digitados os títulos de cada coluna.
- Na primeira coluna, e somente na primeira coluna, deverão ser digitados os títulos de cada linha.

The diagram shows an Excel spreadsheet with columns A through F and rows 1 through 12. A rounded rectangle highlights the first row (row 1), with a label 'Títulos das colunas' pointing to it. Another rounded rectangle highlights the first column (column A), with a label 'Títulos das linhas' pointing to it. The intersection of these two rectangles is the cell A1.

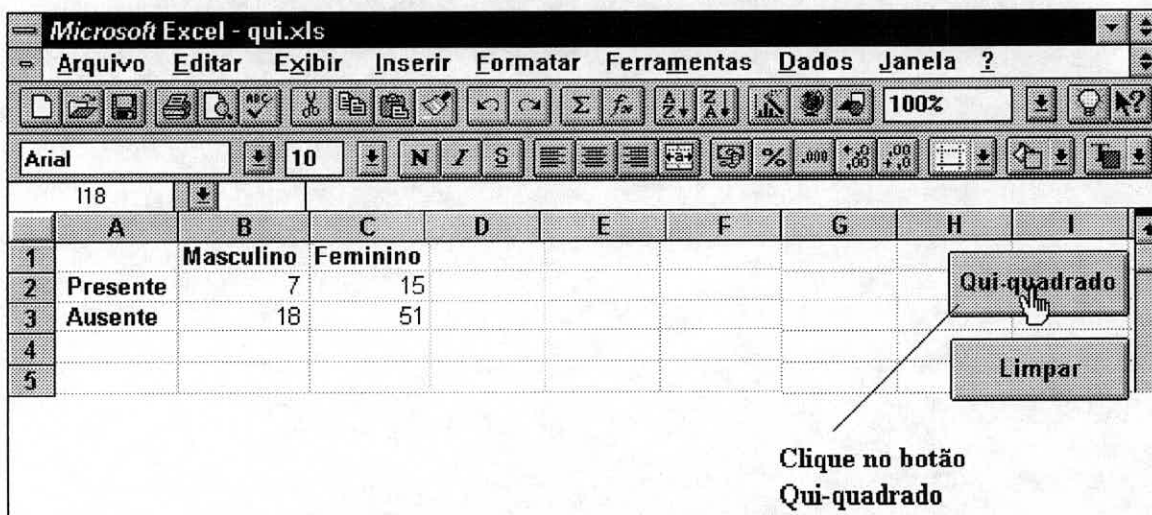
	A	B	C	D	E	F
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						

O problema que vamos resolver é o seguinte: A depressão que até poucas décadas era considerada uma manifestação puramente subjetiva, hoje é aceita como sendo uma patologia provida de base biológica passível de tratamento medicamentoso específico. A literatura mundial refere a uma incidência maior de depressão no sexo feminino. Verificar se existe dependência entre as seguintes variáveis: Depressão (Presente, Ausente) e Sexo (Masculino e Feminino).

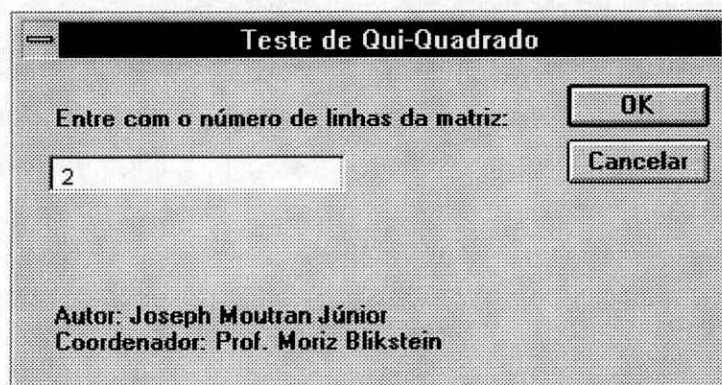
Digite os dados como abaixo:



Após os dados digitados, simplesmente clique no botão **Qui-quadrado**:



Na seguinte tela, digite o número de linhas da matriz:



Teste de Qui-Quadrado

Entre com o número de linhas da matriz:

2

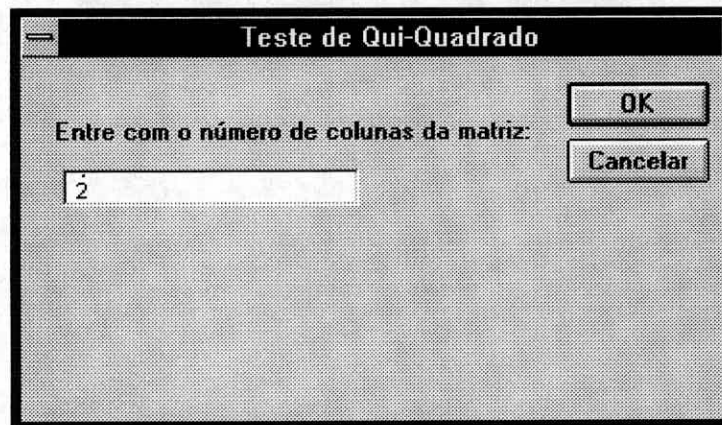
OK

Cancelar

Autor: Joseph Moutran Júnior
Coordenador: Prof. Moriz Blikstein

Pressione o botão OK.

Na próxima tela, digite o número de colunas da matriz de dados:



Teste de Qui-Quadrado

Entre com o número de colunas da matriz:

2

OK

Cancelar

Pressione o botão OK.

Surgirá o resultado do Qui-quadrado:

Microsoft Excel - qui.xls

Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela ?

Arial 10 N I S

A16

	A	B	C	D	E	F	G	H	I
1		Masculino	Feminino	Total					
2	Presente	7	15	22					Qui-quadrado
3	Ausente	18	51	69					
4	Total	25	66	91					Limpar
5									
6									
7	Matriz do qui-quadrado:								
8									
9		Masculino	Feminino						
10	Presente	0,15	0,06						
11	Ausente	0,05	0,02						
12									
13	O valor do qui-quadrado é:				0,27				
14	O Número de Graus de Liberdade é:				1				
15	O Valor de P é:				60,00%				
16									

Plan1 / Colunas / Linhas / Module1 /

Pronto Soma=0 NUM

EXERCÍCIO DE REGRESSÃO

Coordenador: Joseph Moutran Júnior

Orientador: Prof. Moriz Blikstein

EXERCÍCIO DE REGRESSÃO

Esta apostila mostra os recursos básicos da planilha Excel para realizar uma regressão linear simples e também uma regressão linear múltipla.

1 - ENTRANDO NO EXCEL

Ao ligar o computador, digite **WIN** para entrar no ambiente Windows.

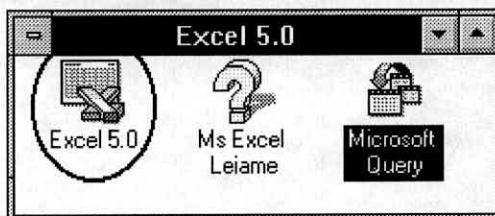
Quando estiver dentro do Windows, localize o ícone **Excel 5.0**.

Leve o indicador do mouse até ele e dê dois cliques rapidamente.

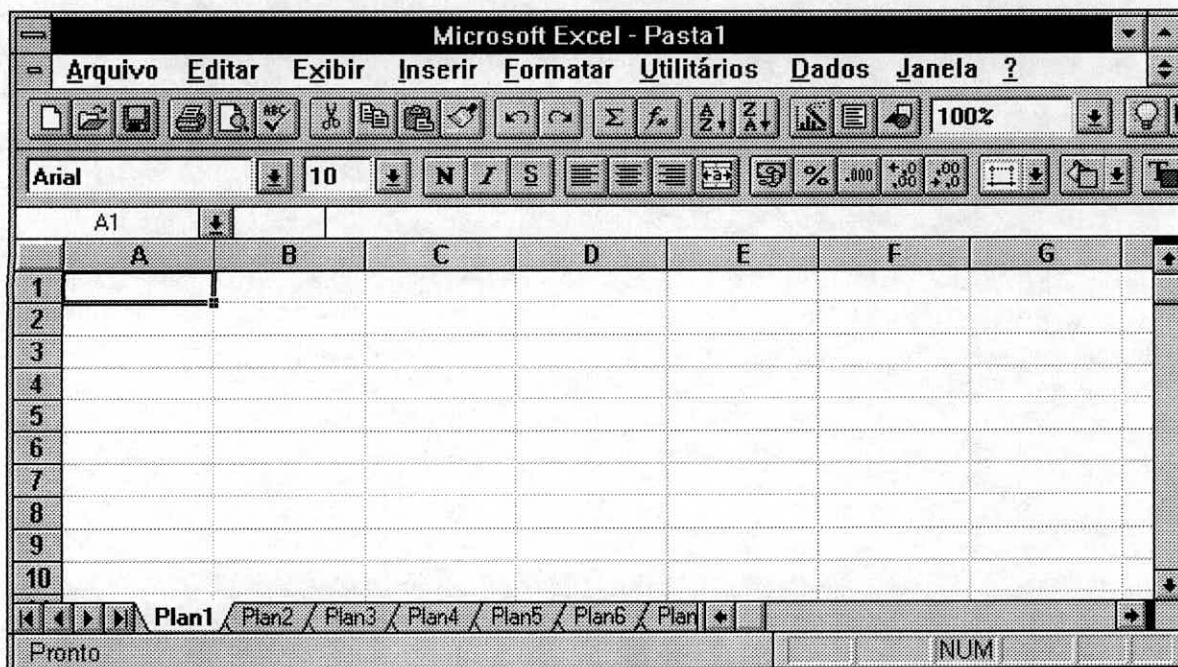
O ícone se transformará numa janela como a mostrada abaixo:



Excel 5.0



Leve o indicador do mouse até o ícone Microsoft Excel, assinalado na figura, e dê um duplo clique sobre ele. Surgirá a tela do Excel e você já estará pronto para digitar os seus dados.



2 - DIGITANDO OS DADOS EM SUA PLANILHA

Iremos realizar um exercício de regressão linear simples. Este exercício consiste em fazer uma regressão tendo como variável dependente (**Y**) a frequência de acidentes para cada 100 trabalhadores.

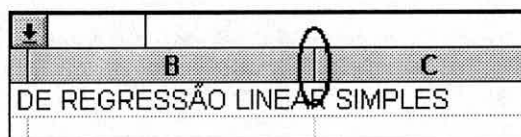
Como variável independente (**X**) teremos o tempo médio de serviço (em meses).

Para isso, digite os dados como mostrado na planilha abaixo:

	A	B	C
1	EXERCÍCIO DE REGRESSÃO LINEAR SIMPLES		
2			
3	SETOR	ACIDENTES (1000 trab.)	T. SERV. (meses)
4		Y	X
5	A	37	3,2
6	B	32	8,1
7	C	26	7,9
8	D	20	8
9	E	20	10,2
10	F	16	12,1
11	G	15	9,3
12	H	12	11,2

DICAS:

- Para ajustar a largura das colunas, leve o indicador do mouse até a borda direita da primeira célula da coluna e dê um duplo clique. Por exemplo, se quisermos ajustar a largura da coluna B, devemos levar o mouse até a posição determinada na figura abaixo e dar um duplo clique.



- Para centralizar o conteúdo de uma determinada célula:
Selecione com o mouse o intervalo desejado

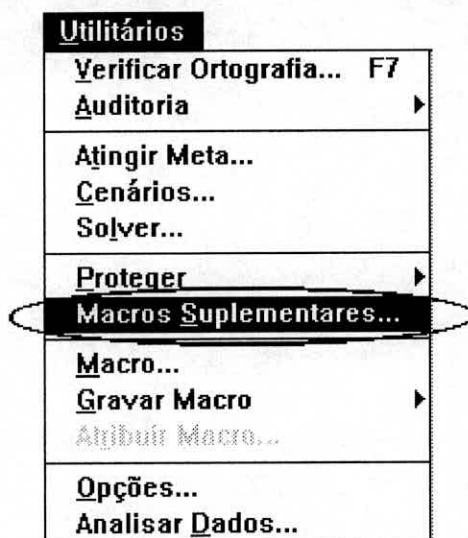
Clique com o mouse no seguinte botão

3 - EXECUTANDO UMA REGRESSÃO LINEAR

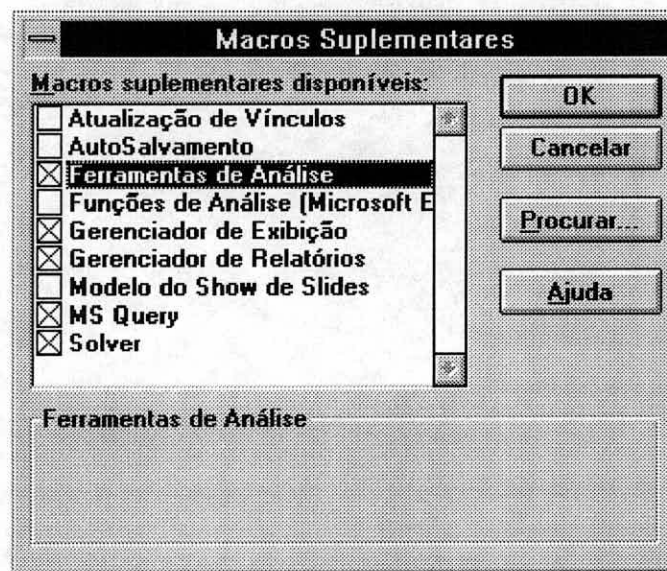
Em primeiro lugar devemos carregar o pacote estatístico do Excel. Para isso:

1. Leve o cursor até a barra de menus (segunda barra na tela de cima para baixo).
2. Clique no comando **Utilitários**

Surgirá um menu com várias opções. Clique na opção **Macros Suplementares...**



Neste instante aparecerá uma nova janela como indicado na figura abaixo:



Se a opção Ferramentas de Análise não tiver um X ao seu lado, clique em cima dela.

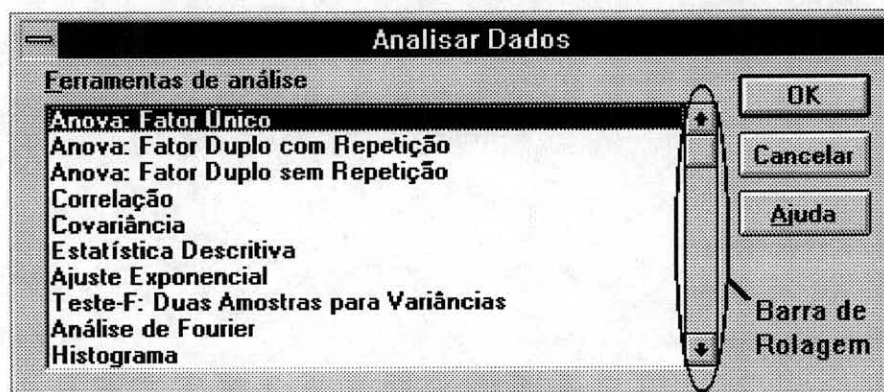
Clique o botão OK

Com isto as Ferramentas de Análise estarão prontas para serem usadas.

Vamos agora realizar a regressão propriamente dita. Para isso:

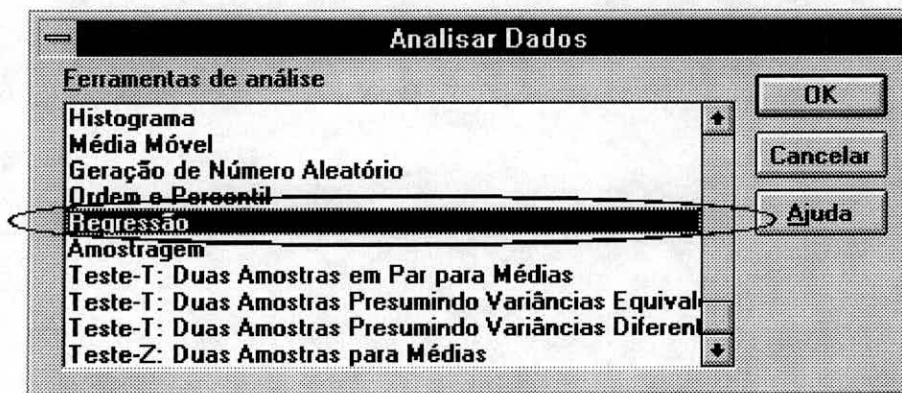
1. Leve o cursor até o menu **Utilitários**
2. Escolha a opção **Analisar Dados**

Surge a seguinte caixa de diálogo:



Nela poderemos escolher entre vários tipos de ferramentas estatísticas. A ferramenta que usaremos não aparece na tela. Portanto teremos que usar a barra de rolagem até a opção

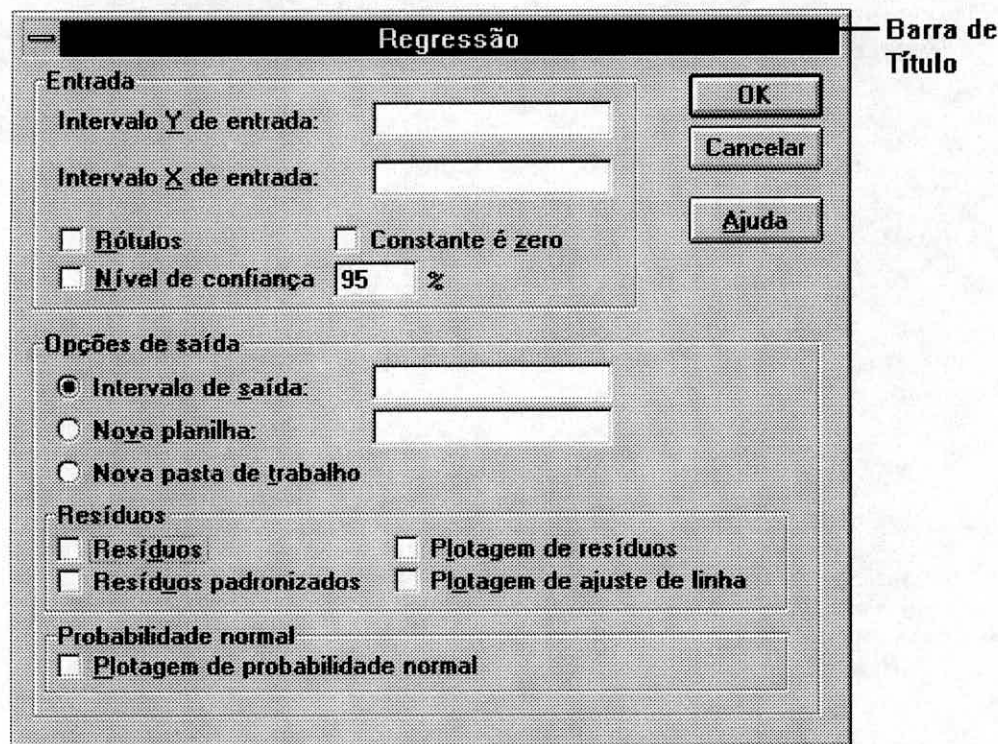
Regressão aparecer na tela. Para isso, na barra de rolagem, clique na seta que aponta para baixo.



Quando a opção **Regressão** aparecer:

1. Dê um clique nela.
2. Clique no botão **OK**

Surge nova Caixa de Diálogo, como mostra a figura abaixo:



Para realizarmos a regressão, precisaremos visualizar esta janela e os dados na planilha ao mesmo tempo. Podemos mudar esta janela de lugar através de sua barra de título (mostrada na figura acima). Para isso:

1. Leve o mouse até qualquer ponto da barra de título.
2. Clique nela e com o botão do mouse pressionado, arraste o mouse para a direita até visualizar os dados.

A primeira informação que aparece na caixa de diálogo é **Intervalo Y de Entrada**. Nesta caixa teremos que digitar o intervalo da planilha onde se encontram os dados da variável Y. Este intervalo vai da célula **B5** até a célula **B12**.

Podemos digitar **B5:B12** ou simplesmente selecionar com o mouse este intervalo. Para isso:

3. Clique na caixa **Intervalo Y de Entrada** e digite **B5:B12**

O próximo passo é definir o Intervalo X de Entrada.

4. Clique na caixa em branco ao lado de **Intervalo X de Entrada** e digite **C5:C12**.

Definidos os intervalos das variáveis, vamos agora definir o nível de confiança para a regressão. Na janela de regressão, é sugerido um nível de confiança de 95%. Vamos trabalhar com este nível. Para isto:

5. Clique no quadrinho em branco ao lado da mensagem **Nível de confiança**.

Vamos definir o intervalo onde a regressão será colocada. O intervalo deverá ser qualquer parte da planilha que esteja em branco:

6. Clique no círculo em branco ao lado da mensagem **Intervalo de Saída**
7. Agora, clique na caixa em branco ao lado da mensagem **Intervalo de Saída**
8. Clique na célula **A17**
9. Por último, vamos solicitar que o Excel calcule os resíduos da regressão. Clique na opção **Resíduos**

Sua janela de regressão deverá ter a seguinte aparência:

Regressão

Entrada

Intervalo Y de entrada:

Intervalo X de entrada:

☐ Rótulos ☐ Constante é zero

☒ Nível de confiança %

Opções de saída

☒ Intervalo de saída:

☐ Nova planilha:

☐ Nova pasta de trabalho

Resíduos

☒ Resíduos ☐ Plotagem de resíduos

☐ Resíduos padronizados ☐ Plotagem de ajuste de linha

Probabilidade normal

☐ Plotagem de probabilidade normal

OK Cancelar Ajuda

Se a janela estiver de acordo com a figura acima, pressione **OK**.

A regressão será calculada e os resultados aparecerão a partir da célula **A17**.

	A	B	C	D	E	F	G
17	RESUMO DOS RESULTADOS						
18							
19	<i>Estatística de regressão</i>						
20	R múltiplo	0,85739967					
21	R-Quadrado	0,735134195					
22	R-quadrado ajustado	0,690989894					
23	Erro padrão	4,852935661					
24	Observações	8					
25							
26	ANOVA						
27		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>	
28	Regressão	1	392,1940928	392,19409	16,652981	0,00649618	
29	Resíduo	6	141,3059072	23,550985			
30	Total	7	533,5				
31							
32		<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>
33	Interseção	46,2478903	6,125862777	7,5496125	0,0002802	31,25843311	61,23734748
34	Variável X 1	-2,74261603	0,67207698	-4,0808064	0,0064962	-4,387130364	-1,098101704
35							
36							
37							
38	RESULTADOS DE RESÍDUOS						
39							
40	<i>Observação</i>	<i>Y previsto</i>	<i>Resíduos</i>				
41	1	37,47151899	-0,471518987				
42	2	24,03270042	7,967299578				
43	3	24,58122363	1,418776371				
44	4	24,30696203	-4,306962025				
45	5	18,27320675	1,726793249				
46	6	13,06223629	2,937763713				
47	7	20,74156118	-5,741561181				
48	8	15,53059072	-3,530590717				

4 - REGRESSÃO LINEAR MÚLTIPLA

Vamos agora realizar uma regressão linear múltipla.

Para tanto, em primeiro lugar, vamos definir uma nova coluna que será usada para a segunda variável independente que é **X2**, na coluna **D**. Esta variável é o investimento mensal em treinamento (R\$ por trabalhador).

1. Na célula **A1**, digite EXERCÍCIO DE REGRESSÃO LINEAR MÚLTIPLA
2. Na célula **C4**, digite **X1**
3. Na coluna **D**, digite os dados de acordo com a figura abaixo:

	A	B	C	D
1	EXERCÍCIO DE REGRESSÃO LINEAR MÚLTIPLA			
2				
3	SETOR	ACIDENTES (1000 trab.)	T. SERV. (meses)	TREINAMENTO (\$)
4		Y	X1	X2
5	A	37	3,2	200
6	B	32	8,1	230
7	C	26	7,9	270
8	D	20	8,0	290
9	E	20	10,2	320
10	F	16	12,1	360
11	G	15	9,3	400
12	H	12	11,2	410

4. Clique no menu **Utilitários** e escolha a opção **Analisar Dados**
5. Como já feito anteriormente, escolha a opção **Regressão**
6. Clique em **OK**

Surgirá novamente a janela da regressão.

Poderemos aproveitar quase todos os dados da regressão linear simples feita anteriormente.

A única modificação que faremos é redefinir o **Intervalo X de Entrada**

1. Pressione uma vez a tecla **TAB** para mover o cursor até a opção **Intervalo X de Entrada**
2. Digite **C5:D12**

Sua janela de regressão deverá ficar como abaixo:

Regressão

Entrada

Intervalo Y de entrada: B5:B12

Intervalo X de entrada: C5:D12

☐ Rótulos ☐ Constante é zero

☒ Nível de confiança 95 %

Opções de saída

☒ Intervalo de saída: A17

☐ Nova planilha:

☐ Nova pasta de trabalho

Resíduos

☒ Resíduos ☐ Plotagem de resíduos

☐ Resíduos padronizados ☐ Plotagem de ajuste de linha

Probabilidade normal

☐ Plotagem de probabilidade normal

OK Cancelar Ajuda

Clique no botão **OK**.

Será recalculada a regressão e o resultado aparecerá a partir da célula **A17**

	A	B	C	D	E	F	G
16							
17	RESUMO DOS RESULTADOS						
18							
19	<i>Estatística de regressão</i>						
20	R múltiplo	0,972782303					
21	R-Quadrado	0,946305408					
22	R-quadrado ajustado	0,924827571					
23	Erro padrão	2,393577437					
24	Observações	8					
25							
26	ANOVA						
27		<i>gl</i>	<i>SQ</i>	<i>MQ</i>	<i>F</i>	<i>F de significação</i>	
28	Regressão	2	504,8539353	252,4269676	44,0596239	0,000668077	
29	Resíduo	5	28,64606473	5,729212946			
30	Total	7	533,5				
31							
32		<i>Coefficientes</i>	<i>Erro padrão</i>	<i>Stat t</i>	<i>valor-P</i>	<i>95% inferiores</i>	<i>95% superiores</i>
33	Interseção	56,12597038	3,753811825	14,9517272	2,42245E-05	46,47650566	65,7754351
34	Variável X 1	-0,708228831	0,565997103	-1,25129409	0,266182003	-2,163168327	0,746710665
35	Variável X 2	-0,089286994	0,020134976	-4,43442269	0,006800038	-0,141045513	-0,037528475
36							
37							
38							
39	RESULTADOS DE RESÍDUOS						
40							
41	<i>Observação</i>	<i>Y previsto</i>	<i>Resíduos</i>				
42	1	36,00223934	0,997760659				
43	2	29,85330825	2,146691747				
44	3	26,42347426	-0,423474263				
45	4	24,5669115	-4,566911501				
46	5	20,33019826	-0,330198256				
47	6	15,41308372	0,586916279				
48	7	13,82464469	1,175355309				
49	8	11,58613997	0,413860026				

Autor: Joseph Moutran Júnior
Orientador: Prof. Moriz Blikstein

MÉTODO DE HOLT-WINTERS USANDO O EXCEL

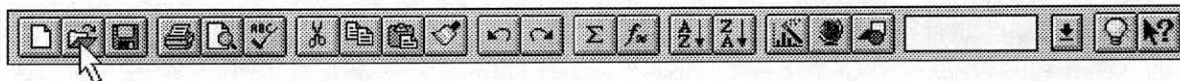
Vamos abrir o programa Excel. Para isso dê dois clique no ícone mostrado abaixo:



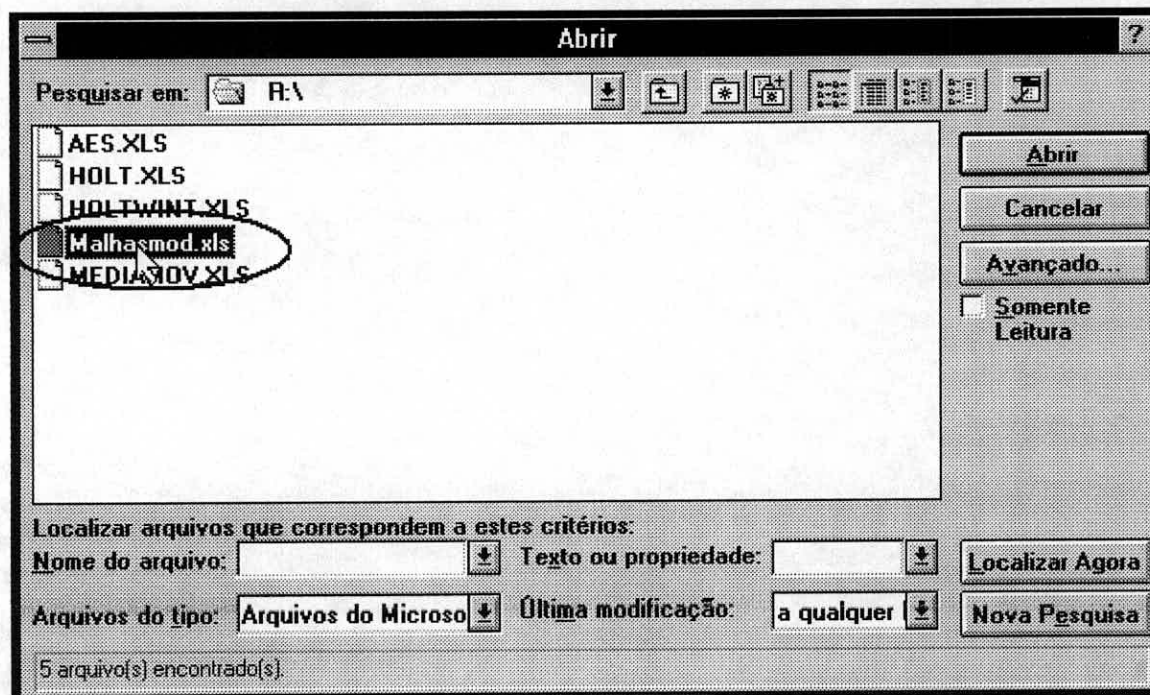
Na janela que se abre, dê um duplo clique no ícone do Microsoft Excel mostrado abaixo:



Em primeiro lugar, abra o arquivo **HOLTWINT.XLS** que se encontra na unidade **P**. Para isso, clique com o mouse no botão **Abrir**:



Selecione a unidade **A:**, o arquivo **Malhasmod.xls**



Teremos a seguinte planilha em branco:

	A	B	C	D	E	F	G	H	I	J	K
1	A:		B:		C:		EQM:		EPAM:		
2											
3	t	ano	sazon	X (t)	Xbar (t)	T (t)	S (t)	F (t)	E (t)	EPA	
4	1										
5	2										
6	3										
7	4										
8	5										
9	6										
10	7										
11	8										
12	9										
13	10										
14	11										
15	12										
16	13										
17	14										
18	15										
19	16										

Base / Linhas / Sazonalidade / SLinha / Módulo1 /

Pronto Soma=0 NUM

Vamos agora preencher a planilha:

Iremos usar como modelo o seguinte exercício de autoria do Prof. Jorge Motta.:

A empresa Malhasmoda Indústria e Comércio de Malhas Ltda. Fabrica e comercializa malhas para adultos e crianças desde sua fundação em 1957. Em dezembro de 1996, sua diretoria comercial estava empenhada em estimar as vendas de janeiro de 1997. Os dados apresentados a seguir abrangem o período compreendido entre janeiro de 1993 e dezembro de 1996. Os valores são expressos em US\$ 1.000 dólares. Os dados seguem abaixo:

<u>MÊS</u>	<u>ANO</u>			
	<u>1993</u>	<u>1994</u>	<u>1995</u>	<u>1996</u>
janeiro	4674	3811	5419	5157
fevereiro	5364	4081	6106	6408
março	5589	4772	6401	6936
abril	6315	5361	7343	7581
maio	6698	6650	7515	8536
junho	7098	7096	7697	8499
julho	5971	5793	6846	7004
agosto	6212	6430	7308	7958
setembro	6436	6860	7620	8217
outubro	5550	7082	6479	7682
novembro	5502	6812	4413	7124
dezembro	<u>3592</u>	<u>4589</u>	<u>3467</u>	<u>4771</u>
Total	69001	69337	76614	85873

Um exame atento do quadro apresentado acima revela que as vendas da Malhasmoda Indústria e Comércio de Malhas Ltda. Tendem a aumentar no período de quatro anos, porém variam enormemente dentro de cada ano. A direção geral das vendas de um mês para o seguinte relaciona-se com variações ocorridas durante o mesmo período em outros anos. Por conseguinte, cada mês dentro do ano parece conter influência sazonal. As vendas são registradas em preços constantes, uma vez que alterações de preço influenciaram inevitavelmente a análise.

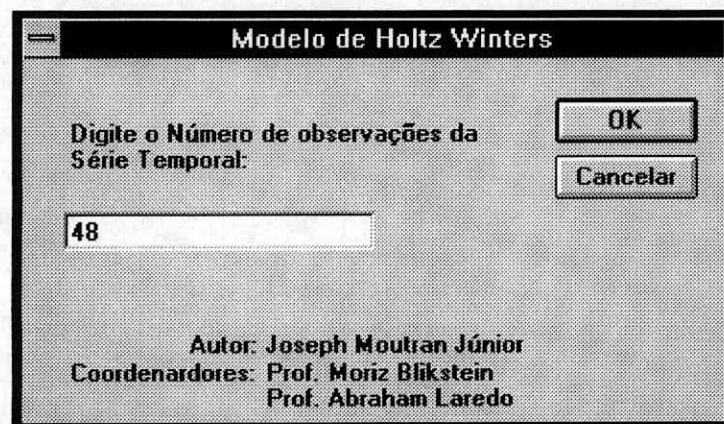
(Dados colhidos pelo Professor Jorge Motta).

Em primeiro lugar digite os valores da Série na coluna D.

Obs: Para o programa funcionar, não é necessário digitar os anos. Basta digitar os dados originais e os períodos (no exemplo os meses).

Agora clique o botão **Execute**

Surge a seguinte Caixa de Diálogo:



Modelo de Holtz Winters

Digite o Número de observações da Série Temporal:

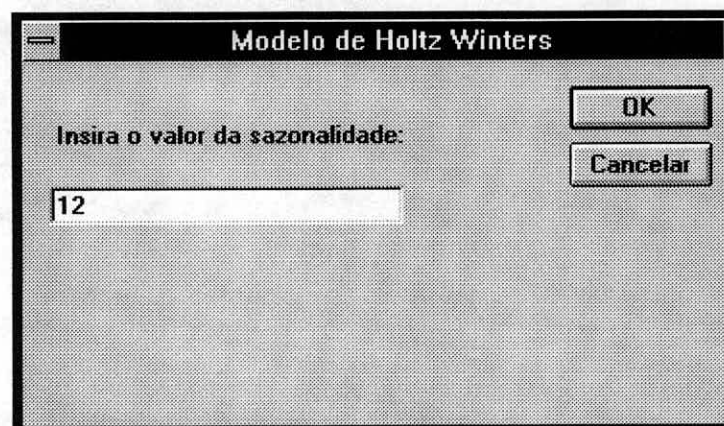
48

OK
Cancelar

Autor: Joseph Moutran Júnior
Coordenadores: Prof. Moriz Blikstein
Prof. Abraham Laredo

Digite **48** e pressione o botão **OK**.

A próxima Caixa de Diálogo que aparece é:



Modelo de Holtz Winters

Insira o valor da sazonalidade:

12

OK
Cancelar

Digite **12** que é o valor da sazonalidade no exercício e pressione **OK**.

O resultado final é mostrado na figura da próxima página:

Microsoft Excel - Malhasmod.xls

Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela ?

Arial 10 N I S

	A	B	C	D	E	F	G	H	I	J	K
1	A:	0,8	B:	0,95	C:	0,90	EQM:	568236	EPAM:	10,113	
2											
3	t	ano	per	X (t)	Xbar (t)	T (t)	S (t)	F (t)	E (t)	EPA	
4	1	1993	1	4674	5868,85	8,16	0,74	4244,4	429,56	9,1905	
5	2		2	5364	6008,20	14,71	0,83	4832,1	531,88	9,9156	
6	3		3	5589	6060,04	16,57	0,90	5448,4	140,63	2,5163	Execute
7	4		4	6315	6106,78	18,08	1,02	6212,4	102,56	1,624	
8	5		5	6698	6056,13	14,64	1,15	7179,9	-481,9	7,1951	Limpar Resultados
9	6		6	7098	6009,36	11,57	1,23	7566,3	-468,3	6,5982	
10	7		7	5971	5987,19	9,89	1,02	6214	-243	4,069	
11	8		8	6212	5931,33	6,60	1,09	6648,1	-436,1	7,0205	
12	9		9	6436	5867,60	3,08	1,15	6901,9	-465,9	7,2389	
13	10		10	5550	5711,28	-4,89	1,08	6452,2	-902,2	16,256	
14	11		11	5502	5597,43	-10,34	1,06	6030,6	-528,6	9,6083	
15	12		12	3592	5484,54	-15,46	0,70	3874,5	-282,5	7,8634	
16	13	1994	1	3811	5400,13	-18,91	0,74	4067,4	-256,4	6,7279	
17	14		2	4081	5290,44	-23,45	0,82	4441,3	-360,3	8,8282	
18	15		3	4772	5271,20	-23,24	0,90	4710,7	61,319	1,285	
19	16		4	5361	5253,62	-22,96	1,02	5261,4	99,597	1,8578	

Base / Linhas / Sazonalidade / SLinha / Módulo1

Pronto Soma=0 NUM

Na primeira linha teremos o EQM (Erro Quadrático Médio) e o EPA (Erro Percentual Médio Absoluto).

Autor: Joseph Moutran Júnior
Orientador: Prof. Moriz Blikstein

ANÁLISE DE AGRUPAMENTOS USANDO O SYSTAT

Para abrir o programa Systat, em primeiro lugar dê um duplo clique no ícone de grupo mostrado abaixo:



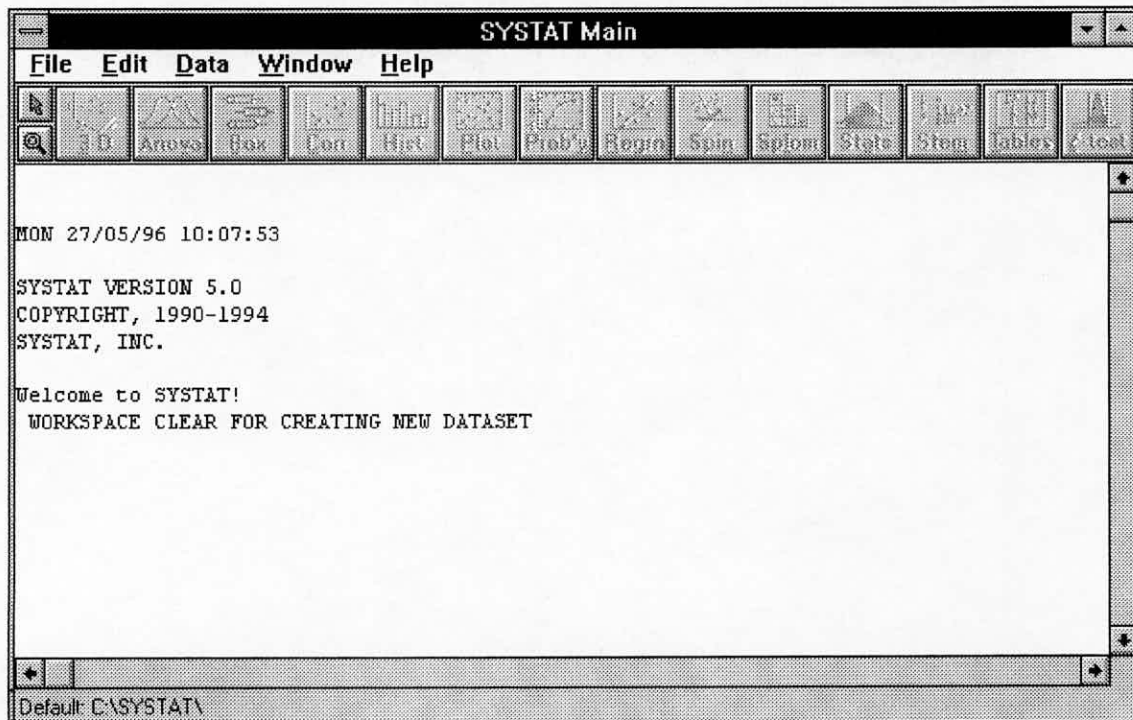
Systat

A janela de nome Systat se abrirá e dentro dela você encontrará o seguinte ícone:



Systat

Dê um duplo clique neste ícone para iniciar o programa. Surgirá a tela do programa:

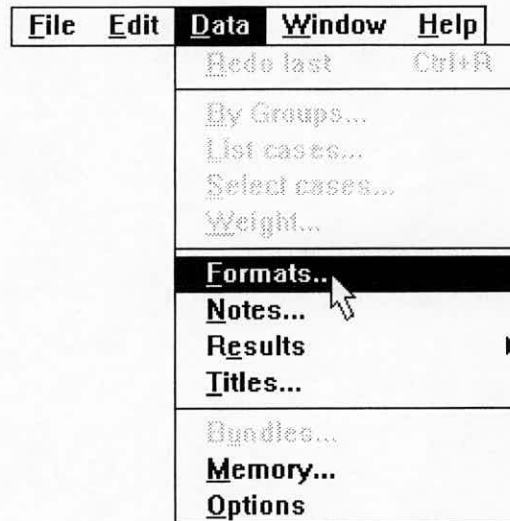


Antes de iniciarmos a análise de agrupamentos, vamos executar um comando para informar ao Systat, o modo como queremos que ele nos forneça os dados. Para isso:

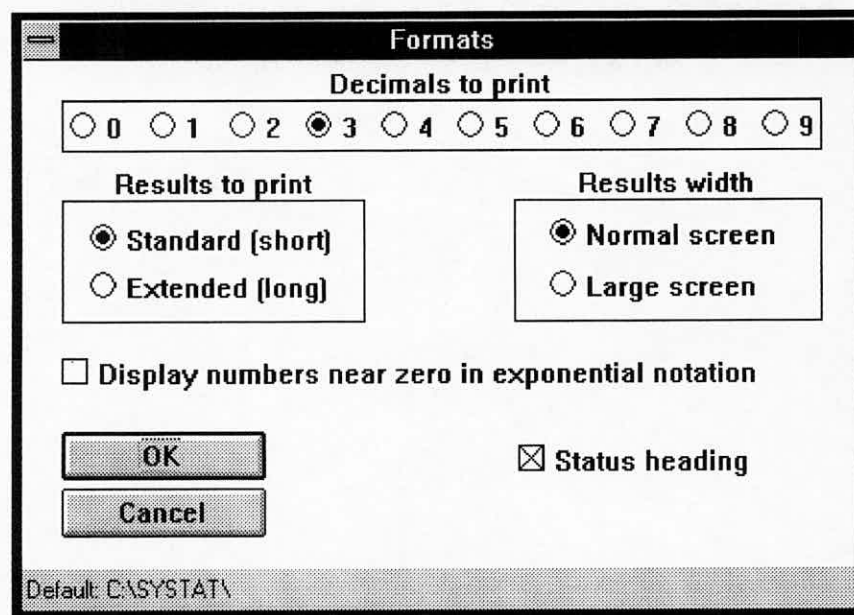
1. Clique com o mouse no menu **Data**.



2. Entre as opções que surgirão, clique com o mouse na opção **Formats**.



Teremos então uma caixa de diálogo como a mostrada abaixo:



Na seção **Results to print**, escolha a opção **Extended (long)** clicando com o mouse nela.

Você também tem a opção de escolher o número de casas decimais com que vai trabalhar na opção **Decimals to print**.

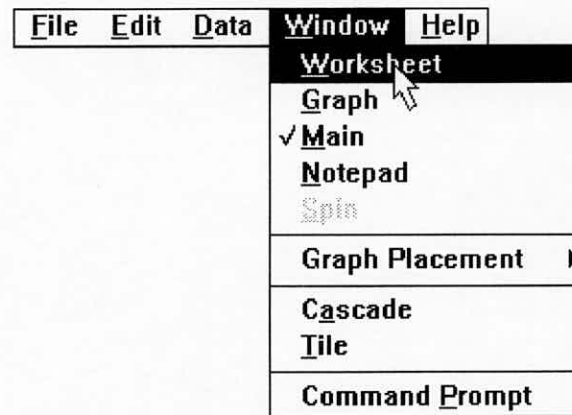
Após realizar as mudanças necessárias, clique no botão **OK**.

Neste instante já estamos prontos para iniciar a análise de agrupamentos. O primeiro passo que devemos realizar é inserir os dados necessários. Faremos isto usando o módulo de planilha do Systat. Para isso:

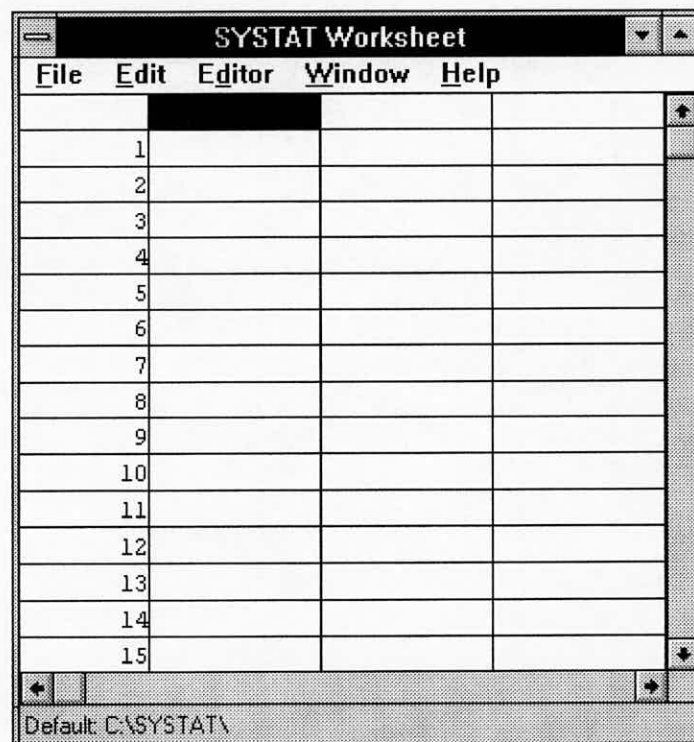
Clique com o mouse no menu **Window**



Surgirão várias opções. Entre elas, escolha a primeira, **Worksheet**.



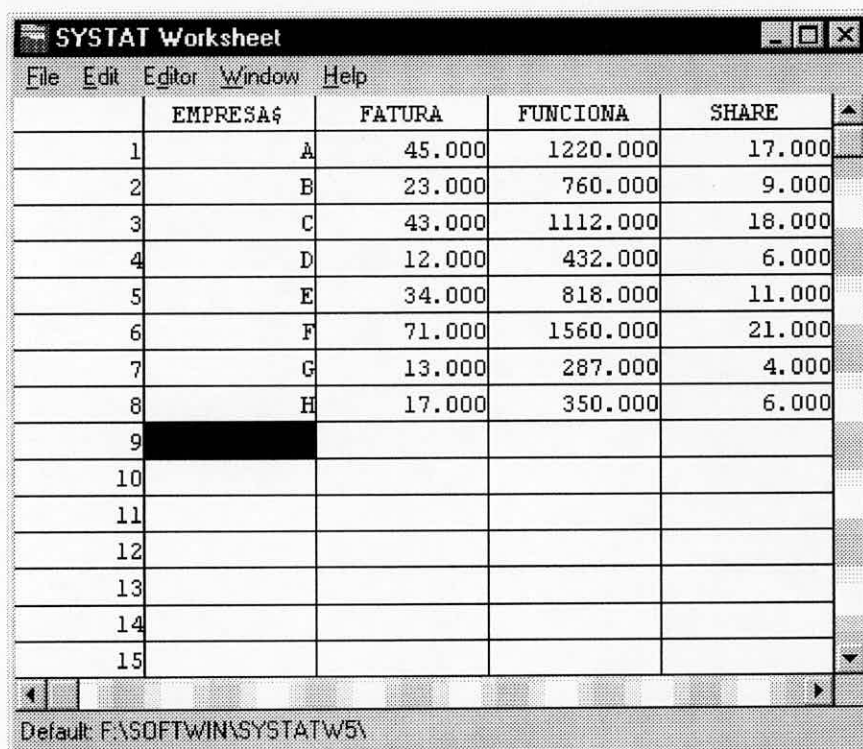
Surgirá, então, a janela da planilha:



Este exercício corresponde a uma análise de agrupamentos de um conjunto de 8 Empresas, de acordo com as seguintes variáveis:

- X1: Faturamento da Empresa, (em milhões de R\$)
 X2: Número de Funcionários
 X3: Market Share

Os dados deverão ser digitados conforme a figura abaixo:



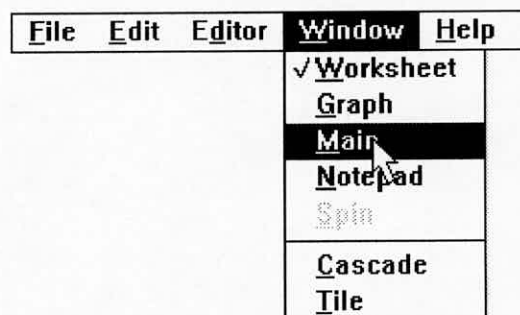
The screenshot shows the SYSTAT Worksheet window with a menu bar (File, Edit, Editor, Window, Help) and a data table. The table has columns for EMPRESA\$, FATURA, FUNCIONA, and SHARE. The data is as follows:

	EMPRESA\$	FATURA	FUNCIONA	SHARE
1	A	45.000	1220.000	17.000
2	B	23.000	760.000	9.000
3	C	43.000	1112.000	18.000
4	D	12.000	432.000	6.000
5	E	34.000	818.000	11.000
6	F	71.000	1560.000	21.000
7	G	13.000	287.000	4.000
8	H	17.000	350.000	6.000
9				
10				
11				
12				
13				
14				
15				

At the bottom of the window, it says: Default: F:\SOFTWIN\SYSTATW5\

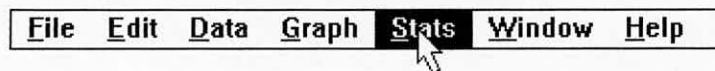
Obs.: Colocamos \$ após o nome da variável **EMPRESA** para indicar para o Systat que se trata de uma variável qualitativa.

Vamos voltar para a tela principal. Para isso clique com o mouse no menu **Window** e selecione **Main**:

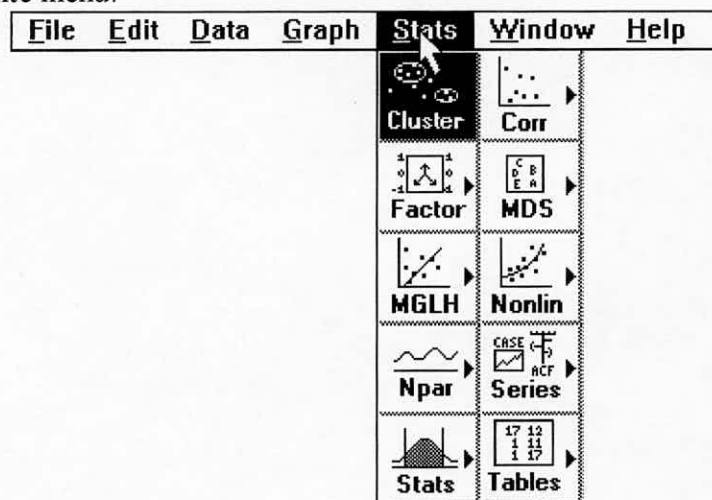


Para realizarmos a análise de agrupamentos (ou análise de Clusters):

1. Na barra de menus, clique no comando **Stats**.



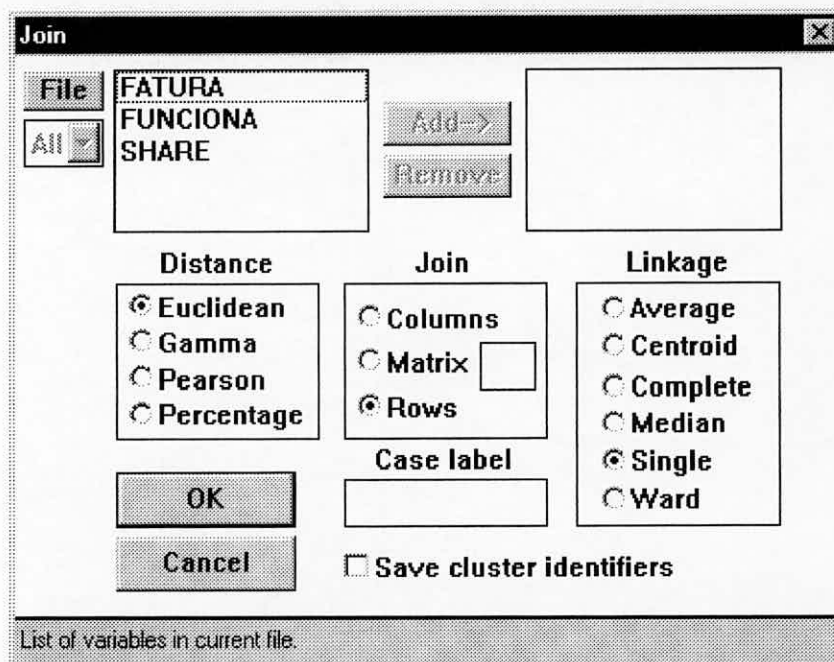
Surgirá o seguinte menu:



2. Clique na opção **Cluster**. Surgirão duas opções como mostrado abaixo:



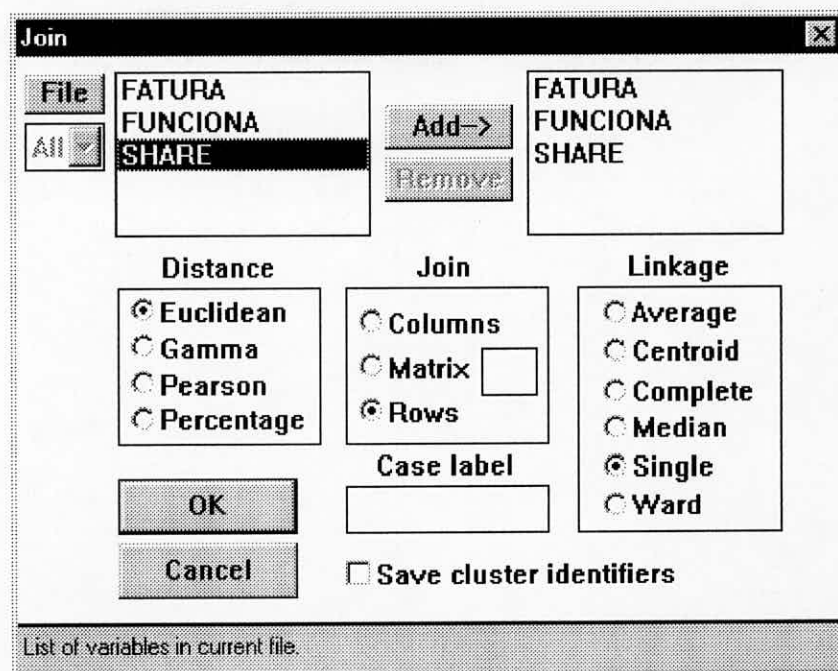
3. Escolha a opção **Join...** Teremos a seguinte caixa de diálogo:



O próximo passo é selecionar as variáveis que serão utilizadas para a análise. Neste exercício iremos utilizar todas as variáveis. Para selecioná-las:

1. No retângulo à esquerda, onde se encontram os nomes das variáveis clique em **Fatura**
2. Clique no botão **Add→**.
3. Repita o procedimento para as variáveis **Funciona** e **Share**.

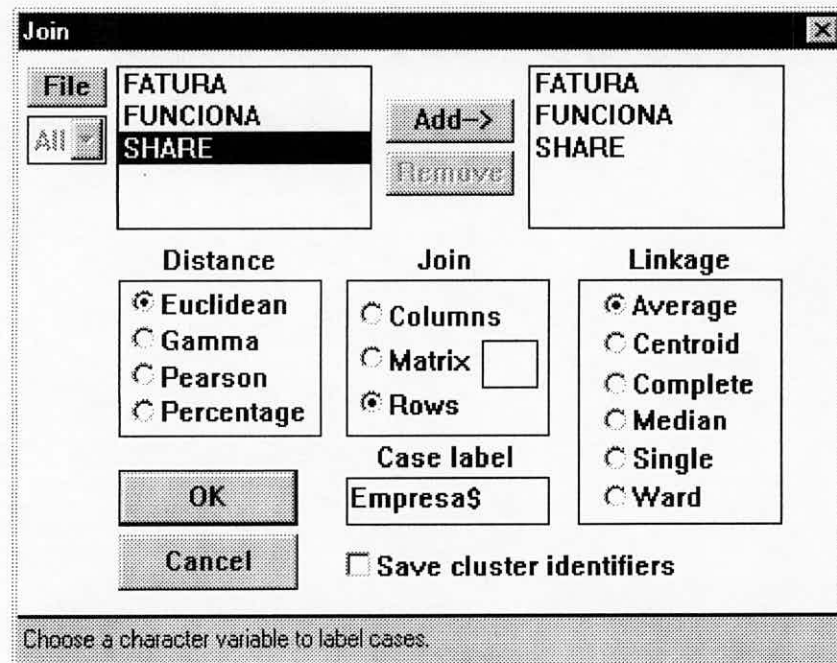
Sua janela deverá estar como a figura abaixo:



Em **Case label** digite o nome da variável nominal. Em nosso caso digite **Empresa\$**

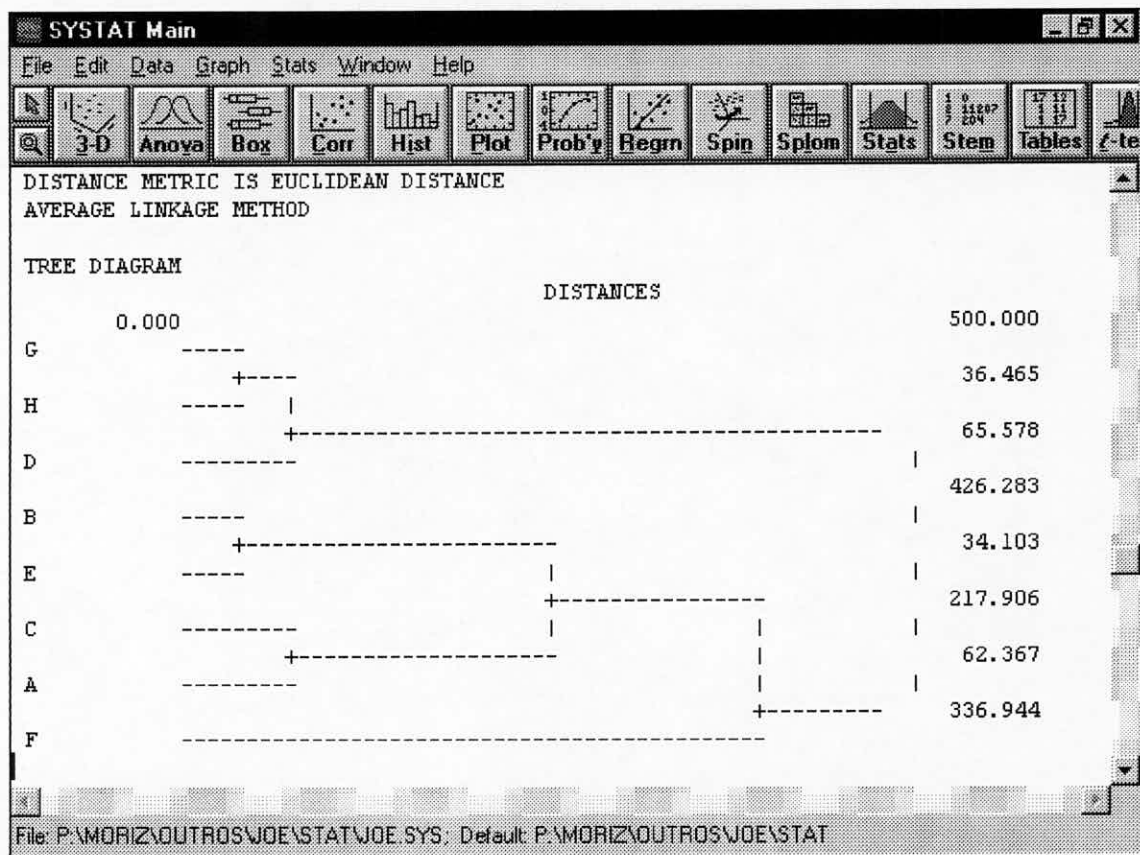
Em **Linkage** escolha a opção **Average**.

Sua caixa de diálogo deverá estar igual à mostrada abaixo:



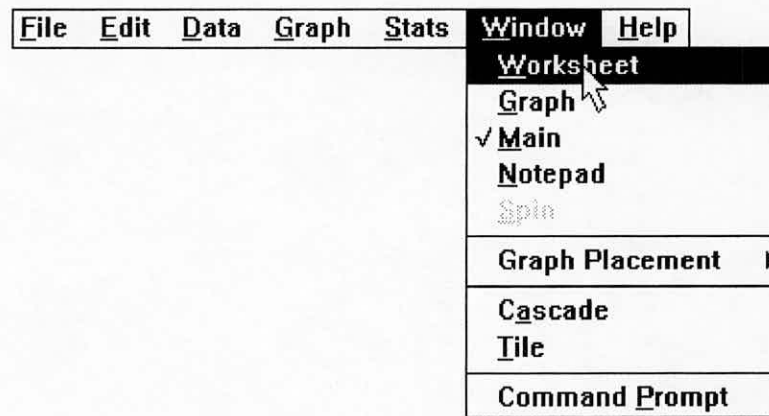
Clique o botão **OK**.

Teremos o dendrograma:



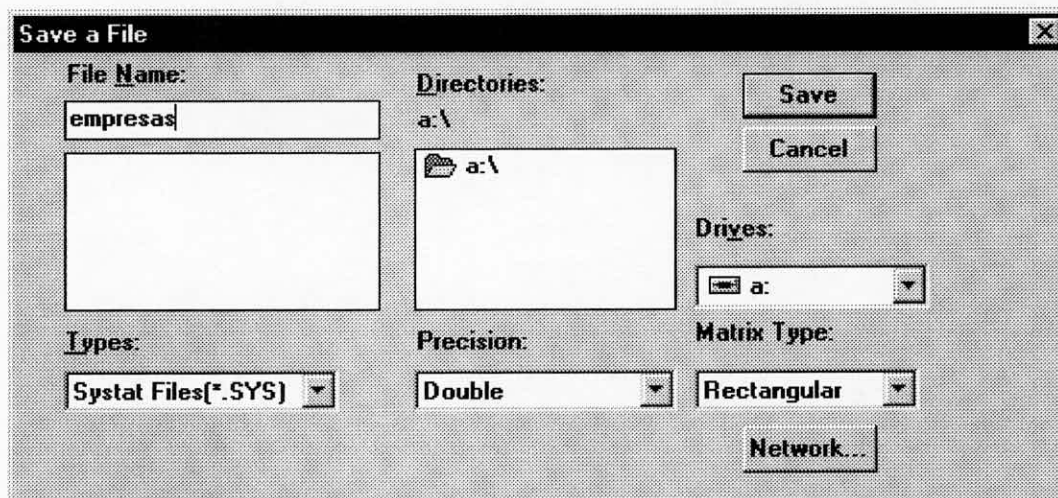
Por último devemos salvar o arquivo com os dados da planilha:

1. Retorne à janela da planilha clicando em **Window** e selecionando **Worksheet**



2. Clique no menu **File** e escolha a opção **Save as**:
3. Em **File Name**, digite **Empresas**.
4. Selecione a unidade (drive) e diretório em que deseja salvar o arquivo.
5. Em drives, selecione a unidade A:

Sua caixa de diálogo deverá ficar como na figura abaixo:



5. Clique o botão **Save**.

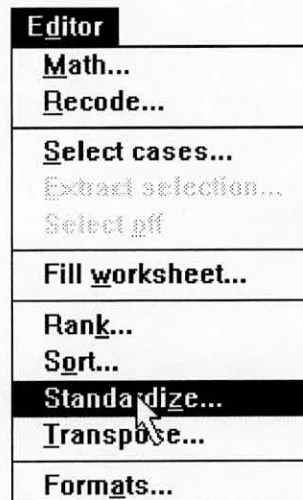
Com isto realizamos o método mais simples de análise de agrupamentos. Entretanto neste caso a estrutura do dendrograma foi influenciada pela escala das variáveis.

Uma das formas de resolvermos este problema é transformando as variáveis antes de utilizá-las para a análise de agrupamentos.

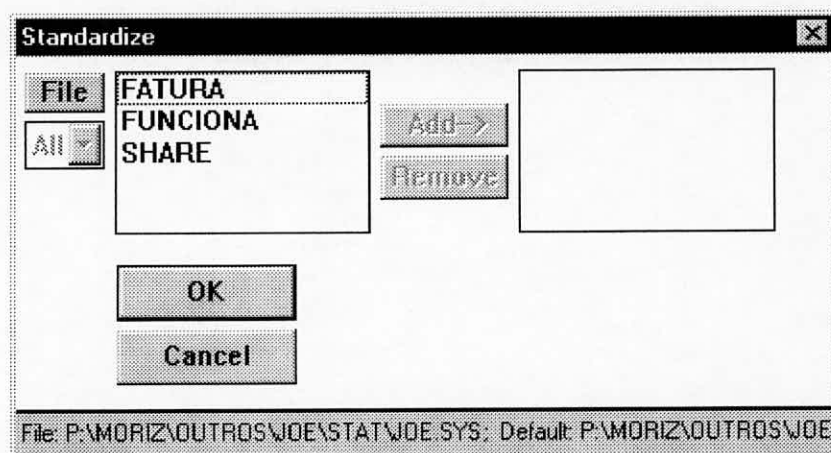
Um primeiro modo de padronização das variáveis é transformá-las de modo que tenham a média zero e a variância igual a 1. Esta padronização é conseguida subtraindo de cada valor observado a média e dividindo a diferença pelo desvio padrão.

Para isso:

1. Retorne à janela da planilha clicando no menu **Window** e escolhendo a opção **Worksheet**.
2. No menu **Editor**, escolha a opção **Standardize**

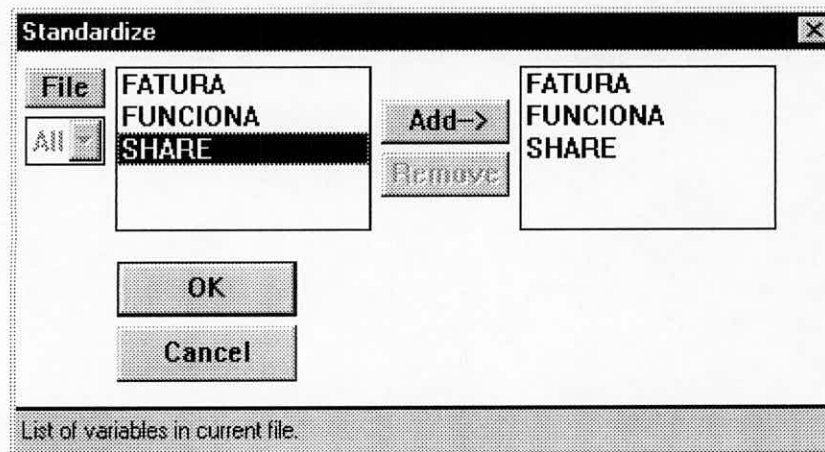


Surge a seguinte caixa de diálogo:



3. Selecione as variáveis que serão padronizadas clicando nelas e posteriormente pressionando o botão 

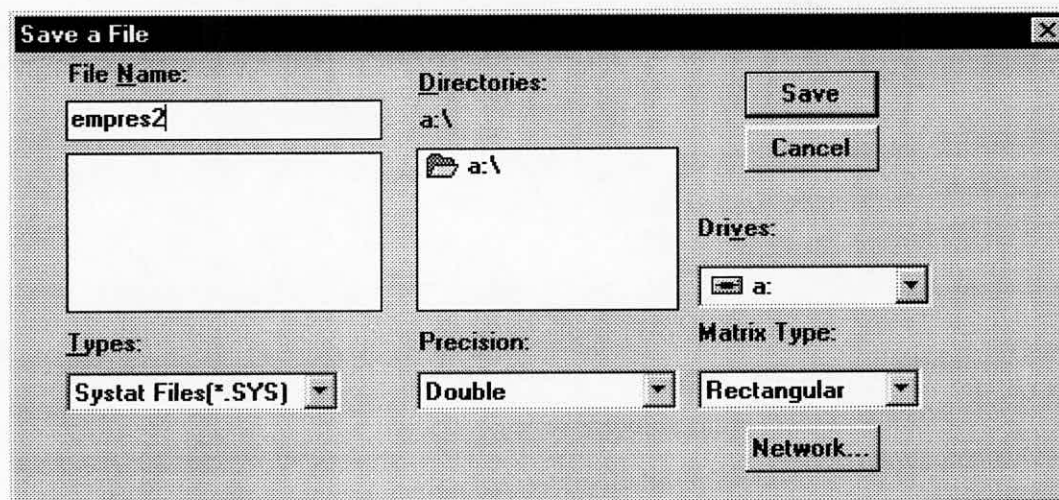
A caixa deverá ficar como abaixo:



4. Clique o botão **OK**.

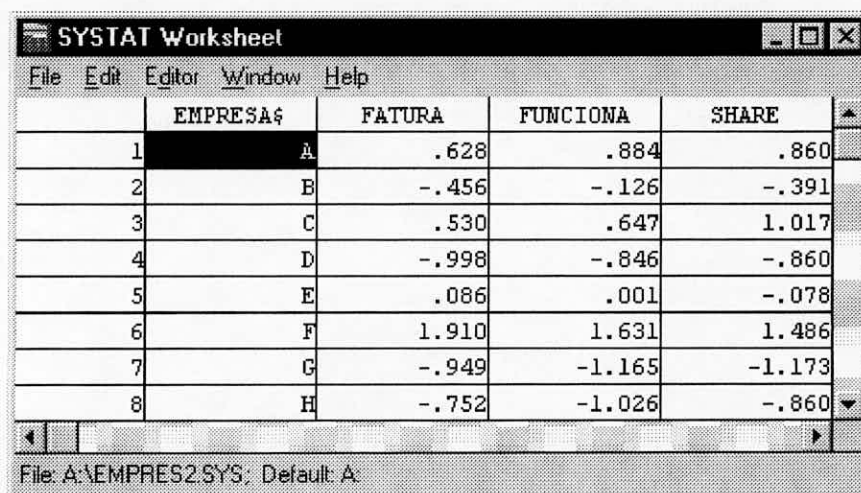
O programa pedirá para você escolher um nome para o novo arquivo que ele irá criar com os dados padronizados. Digite em **File Name** o nome **Empres2**. Escolha o diretório e a unidade em que o arquivo será armazenado.

Este será o aspecto final da caixa de diálogo:



5. Pressione o botão **Save**

A partir deste instante o arquivo **Empresa** com os dados originais foi fechado e em seu lugar estaremos trabalhando com o arquivo **Empres2** conforme figura na outra página:



	EMPRESA	FATURA	FUNCIONA	SHARE
1	A	.628	.884	.860
2	B	-.456	-.126	-.391
3	C	.530	.647	1.017
4	D	-.998	-.846	-.860
5	E	.086	.001	-.078
6	F	1.910	1.631	1.486
7	G	-.949	-1.165	-1.173
8	H	-.752	-1.026	-.860

File: A:\EMPRES2.SYS: Default: A

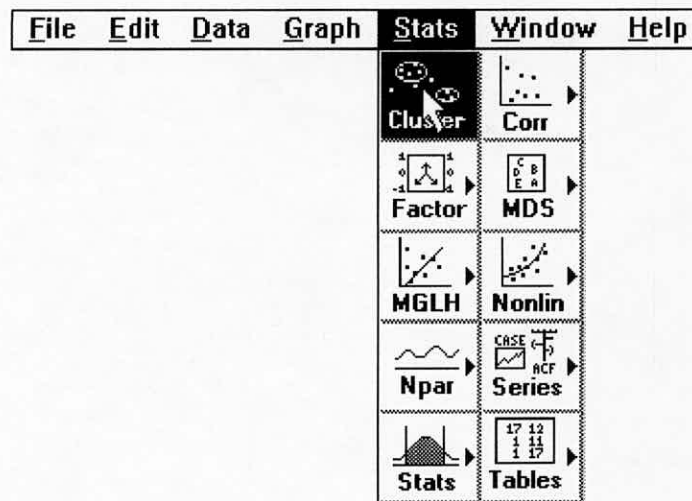
Obs.: Outra forma de transformação é, por exemplo, dividir os valores de cada variável pelo maior valor da mesma. Isto pode ser facilmente feito com a seguinte série de comandos: **Editor, Math**.

Por exemplo, a seguinte variável X1 apresentada abaixo:

X1	X2		X1	X2
4	50		0,4	1,0
10	45		1,0	0,9
7	20		0,7	0,4
3	10		0,3	0,2

Vamos agora realizar a nova análise com os dados previamente padronizados do arquivo PAISES2:

1. Volte para a tela principal clicando em **Window** e escolhendo a opção **Main**
2. No menu **Stats**, selecione a opção **Cluster**

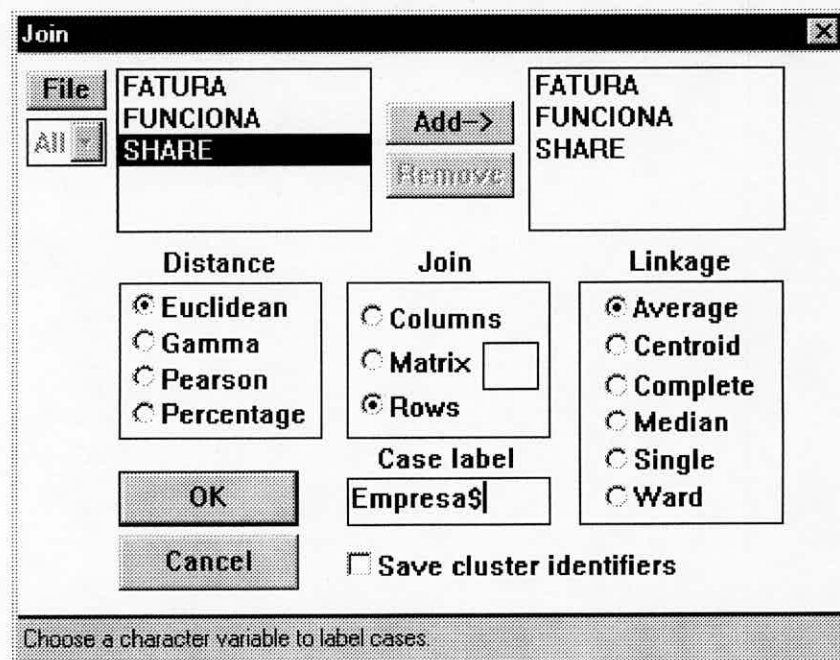


3. Escolha a sub-opção **Join**

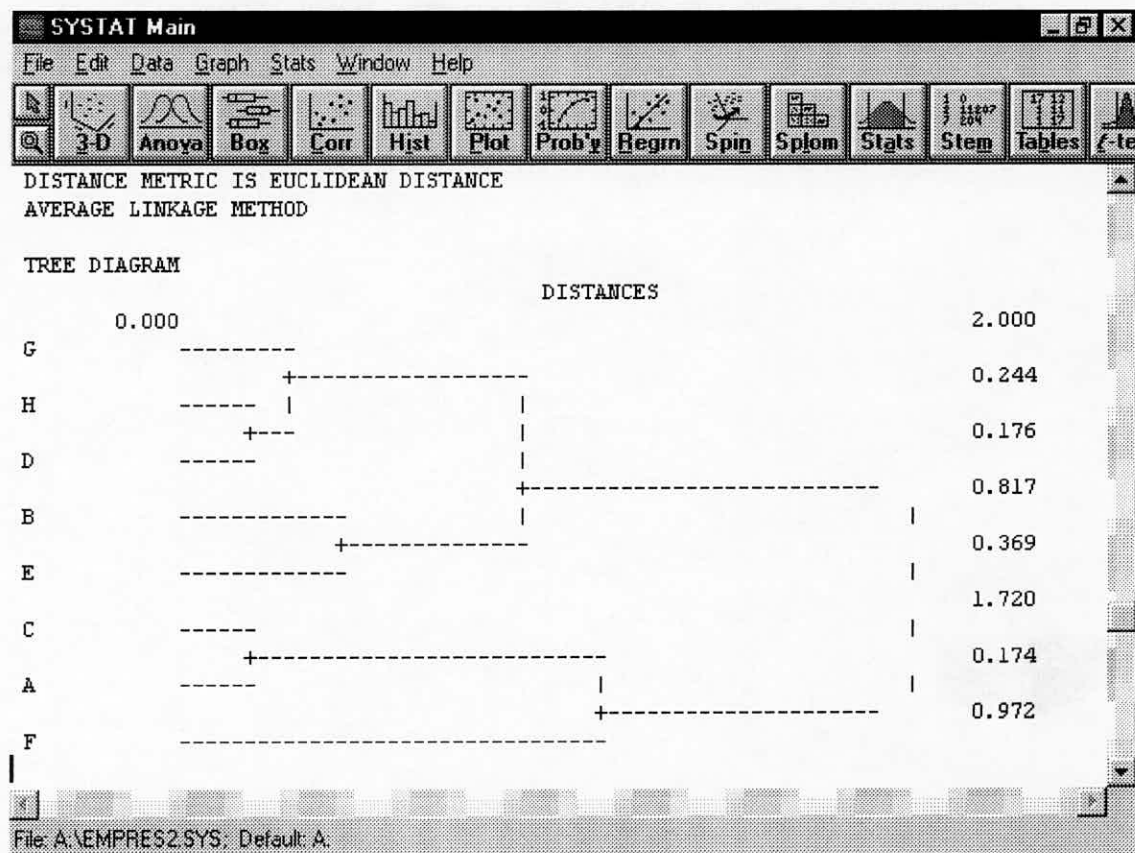
4. Selecione as todas as variáveis clicando nelas e pressionando o botão **Add→**

5. Na caixa **Linkage**, escolha a opção **Average**

6. Em **Case Label**, digite **Empresas\$**



7. Clique em **OK**



Agora temos um dendrograma que não leva em consideração as escalas das variáveis. Repare que este gráfico é diferente do primeiro que levava em consideração as escalas das variáveis.