

A Classificação Hierárquica Implicativa e Coesiva

R. Couturier, A. Bodin, R.Gras

Apresentação

O software CHIC tem por funções essenciais extrair de um conjunto de dados, cruzando sujeitos e variáveis (ou atributos), regras de associação entre variáveis, fornecer um índice de qualidade de associação e de representar uma estruturação das variáveis obtida por meio destas regras. A versão 2.3 de CHIC apresenta melhoras importantes no que diz respeito a convivialidade e as possibilidades de tratamento (por exemplo, é possível tratar tabelas de contingência de tamanho 100*10000). Além de mais, certos erros das versões anteriores foram corrigidos.

Do jeito que está atualmente, o software CHIC já deveria servir de modo importante para a pesquisa assim como para a formação. A divulgação do software CHIC é feita pela ARDM (Associação para a Pesquisa em Didática da Matemática). É possível obtê-lo junto a Regis Gras (endereço na última página). As atualizações do software serão asseguradas gratuitamente durante um ano para toda pessoa ou instituição tendo regularmente obtido uma versão do CHIC. A versão 2.3 só funciona de modo autônomo com Windows 95 ou mais. Um PC 486 (ao menos) é recomendável. Assim sendo, para grandes tabelas de contingência, é preferível ter uma máquina potente (Pentium III com mais de 100 Mo de memória viva). Para o resto, uma familiaridade com Windows é necessária (se não, é recomendável ser acompanhado por um usuário competente). É necessário também ter um mínimo de conhecimento sobre a análise implicativa para poder interpretar corretamente os resultados de diversos tratamentos permitidos pelo CHIC (ver referências e apêndice).

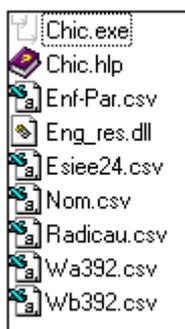
Instalação de CHIC

Para instalar CHIC, criar um repertório (que nomearemos, por exemplo, CHIC) e copiar os elementos do disquete ou dos arquivos fornecidos. Este disquete ou arquivos contém em particular um ou vários arquivos de demonstração que têm a particularidade de ter a extensão *.csv (ver mais adiante a questão do formato dos arquivos). No decorrer desta apresentação, é o arquivo wa392.csv que será usado como arquivo de demonstração. Este arquivo comporta 22 variáveis e 130 indivíduos e é estudado mais completamente em um artigo citado em referência.

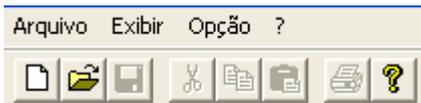
No que segue √ significa CLICAR, → significa “obtemos”.

Iniciando o CHIC

Para iniciar o CHIC, lançar primeiramente Windows, seguido do programa nomeado CHIC.exe.

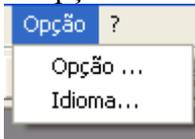


A barra de opções do CHIC aparece assim:

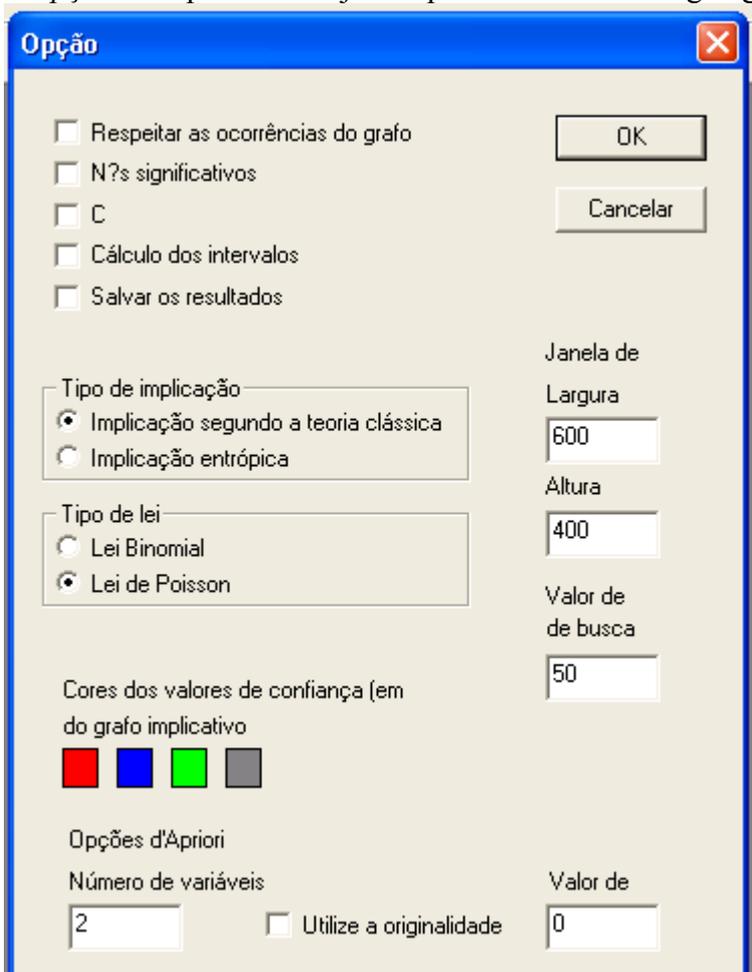


Escolha das opções:

CHIC contém numerosas opções que permitem de configurá-lo como o usuário quiser. Para definir as opções:



√ *Opção* → Aparece uma janela permitindo varias regulagens:



Esses parâmetros são globais. Se eles são modificados, eles agem no próximo tratamento e não no tratamento atual. Mas esta modificação deve ser anterior ao tratamento do arquivo.

- 1 – *Respeite as ocorrências do gráfico!!* **ATENÇÃO ESTA OPÇÃO NÃO TEM, NO MOMENTO, NENHUM EFEITO EM RAZÃO DO USO DO ALGORITMO DE DESENHOS DE GRÁFICO AUTOMÁTICO!!** Esta opção diz respeito ao gráfico implicativo. Se ela é escolhida, um eixo vertical localiza as ocorrências (ou frequências). Será possível mover horizontalmente as variáveis, ou melhor, isto é, mover o nome destas variáveis, (então temos um só grau de liberdade). Senão os elementos do grafo ficarão livres (dois graus de liberdade).

Lembremos que o respeito das ocorrências permite evitar contra-sentidos e outros erros de leitura do gráfico. Mas em contra-partida, o não respeito permite toda reorganização desejada no gráfico.

- **2** – *Nós significativos*: Esta opção, se for assinalada, permite calcular os nós significativos de uma árvore hierárquica, isto é, os que correspondem à uma classificação mais compatível aos valores e à qualidade dos valores de implicação e de coesão.

- **3** – *Cálculo longo*: Esta opção permite fazer com que apareça todo o cálculo intermediário. Ela tem, sobretudo uma função didática e verificadora, mas não é indispensável.

- **4** – *Cálculo dos intervalos*: esta opção diz respeito às variáveis-intervalos, ou seja, as variáveis, como por exemplo, os tamanhos ou os pesos dos indivíduos, cujos valores pertencem a um intervalo de números reais positivos. O usuário está encarregado de definir o número de intervalos, esses últimos não tendo necessariamente a mesma amplitude, e desse fato maximizando a variância interclasse da partição. Em outras palavras, esta partição, calculada pelo programa, “separa” ao melhor, pelo critério de variância, o conjunto dos valores observados de cada uma das variáveis. Em seguida, ela permite calcular as implicações das reuniões destes intervalos.

- **5** – *Salvando os resultados*: **!! ATENÇÃO ESTA OPÇÃO NO MOMENTO NÃO TEM NENHUM EFEITO EM RAZÃO DO USO DO ALGORITMO DE DESENHO DE GRÁFICOS AUTOMÁTICO!!** Esta opção pode se revelar muito útil quando desejamos tratar um volume importante de dados, isto é, quando os tempos dos cálculos são muito importantes. Ela permite ao CHIC de salvar automaticamente o último cálculo efetuado e de usá-lo em um próximo cálculo. Assim o cálculo das correlações é efetuado uma só vez e o resultado desse é reaproveitado se formos escolher a construção de uma árvore de similaridades ou de um gráfico.

- **6** – *Tipo de implicação*: é preciso escolher entre o método clássico e o método entrópico (ver Régis Gras 1997). Este último permite melhor satisfazer ao objetivo de modelagem da inclusão conjuntista na base da teoria da implicação estatística, mas é mais severa no diz respeito à intensidade de implicação.

- **7** – *Tipo de lei*: Duas modelagens do número aleatório de contra-exemplos à implicação ou à regra são propostas: lei binomial e lei de Poisson. O usuário faz sua escolha sabendo que a lei de Poisson é a mais restritiva quanto ao valor da implicação ou da regra (ver Lerman I.C., R. Gras e Rostam H., 1981).

- **8** – *Janela de trabalho*: Esta opção permite escolher a superfície do trabalho. O usuário define uma área de trabalho que ele pode depois modificar durante o cálculo. Assim ele pode escolher uma grande superfície de trabalho no começo que ele reduz em seguida porque certas variáveis não parecem interessantes após o estudo.

- **9** – *Índice de confiança mínimo de pesquisa*: Esta opção tem também por objetivo reduzir o tempo de cálculo quando o número de variáveis é importante. Neste caso, definimos um índice sob o qual não procuraremos mais as implicações entre variáveis. Se o usuário souber de antemão que somente as implicações superiores a um dado índice lhe interessam, esta opção permite a operação.

- **10** – *Opção de conjunção de variáveis (ou item-set em inglês)*: CHIC permite agora calcular conjunções entre variáveis premissas da implicação e isto de maneira automática. Assim, obtemos regras da forma $a \wedge b \wedge c \Rightarrow d$.

A opção “número de variáveis” permite especificar quantas variáveis intervêm nas regras. Se escolhermos 2, teremos então somente as regras da forma $a \Rightarrow b$.

Se escolhermos 3, teremos regras da forma $a \wedge b \Rightarrow c$ e assim em diante. O cálculo das conjunções pode rapidamente gerar um número considerável de regras. Isto tem por consequência tornar delicada a interpretação do gráfico implicativo. Assim, é possível levar em consideração unicamente as regras tendo um caráter original, o que justifica a opção prevista para este efeito, opção essa que permite editar somente as regras superiores ao índice escolhido pelo usuário.

Os quadrados coloridos permitem definir as cores associadas aos diferentes índices do gráfico implicativo.

Escolha da língua

√ *Língua* →

Atenção: esta opção deve ser escolhida antes da abertura do arquivo que será trabalhado.

No momento a versão inglesa é provisória, mas, a termos, podemos prever o espanhol, o italiano, o português...



Ajuda

O menu Ajuda pode ser consultado, mas ele não é muito completo. A ajuda se abre na língua escolhida (por enquanto a ajuda em inglês não foi ainda feita).

√ *Sobre CHIC...* →



Formato dos arquivos

Esta questão é essencial: CHIC tratara somente de arquivos tendo um formato compatível com suas possibilidades de tratamento. Além de mais, os arquivos devem ter um nome do tipo “*.csv” (a extensão sendo obrigatoriamente csv).

O formato que convém é o formato conhecido por Excel sob o nome de CSV. O meio mais seguro de obter um tal arquivo é de digitar os dados ou de transferir um arquivo de dados sobre Excel (Excel 7, 5 ou 4) e depois salvar o arquivo no formato CSV (usando Excel).

O arquivo de dados (uma coluna por variável e uma linha por indivíduo) deve obrigatoriamente ser completado por uma primeira linha contendo os nomes das variáveis e por uma primeira coluna contendo o nome dos indivíduos. Obtemos assim uma tabela do tipo da figura abaixo, na qual a casa

Atenção: O formato csv de um Macintosh não convém. O conveniente seria transferir para um PC o arquivo previamente gravado em formato normal no Mac. Em seguida, no PC, abrir o arquivo com Excel e salvá-lo no formato “separador virgula (csv)”.

Respeitando estas instruções, o usuário poderá passar sem problemas de um universo a outro e de um software a outro. Mesmo assim, a passagem de Mac para PC pode conduzir a dificuldades consideráveis. Todas as casas de valores da tabela devem, obrigatoriamente, ser preenchidas por números compreendidos entre 0 e 1 (menos para as variáveis intervalos). Assim, todo número decimal do intervalo [0;1] pode figurar, no caso das variáveis freqüenciais como por exemplo: 0,80. Se uma casa está vazia ou tem um formato que não convém, o software se bloqueia e assinala o erro, assim como o lugar onde este erro se encontra. No momento, ele não irá mais longe nos cálculos.

Variáveis

CHIC permite realizar a análise de dados a partir de variáveis. Existem vários tipos de variáveis que podem ser tratadas por este software. As variáveis podem ser dos seguintes tipos: binária, modal, freqüencial ou intervalo. Elas podem ser principais ou secundárias (ou suplementares).

- **Variáveis binárias:** toda variável binária, como indica o nome, assume unicamente dois valores, 0 ou 1. Ela significa a **antinomia** entre estes dois valores, como por exemplo, a presença e a ausência, o verdadeiro e o falso, a posse e a não posse, etc. A soma dos elementos de uma coluna representa a ocorrência da variável em questão (número de vezes que ela é satisfeita). A soma dos elementos de uma linha representa o número de variáveis que o sujeito em questão possui ou satisfaz.

- **Variáveis modais e freqüenciais:** na seqüência do nosso trabalho, ampliamos a noção de implicação estatística a outras variáveis além das binárias. É o caso das variáveis modais que são associadas a fenômenos nos quais os valores $a(x)$ (valores atribuídos pelos sujeitos x à variável a) são números pertencentes ao intervalo [0, 1] e que descrevem o grau de pertinência ou de satisfação como na lógica fuzzy. Por exemplo, as modalidades: “concordância plena”, “concordância”, “concordância parcial”, “não concordância”, reveladas em um questionário de opiniões, são levadas em consideração e transformadas em valores de intensidade ordenados, como por exemplo, as modalidades definidas pelos respectivos números: 1, 0.75, 0.50, 0.25 e 0. É também o caso das variáveis freqüenciais, como as porcentagens, que são associadas aos fenômenos em que os valores de $a(x)$ são números reais positivos quaisquer.

J. B. Lagrange (1998) deu, nestes casos, novos índices que coincidem com os índices de implicação entre variáveis binárias.

- **Variáveis quantitativas ou efetivas:** estas variáveis, a valores reais positivos, descrevem situações nas quais aos elementos x de E faz-se corresponder uma quantidade $a(x)$, por exemplo, o número de vezes que a variável a é observada em x . Para tratar essas variáveis, basta, com a ajuda de uma planilha, transformar os valores observados, por uma normalização utilizando o valor máximo observado, em valores do intervalo [0, 1]. A fórmula que calcula o índice de implicação entre variáveis binárias é então utilizada no caso das variáveis freqüenciais.

- **Variáveis- intervalo:** Este tipo de variável é uma extensão do tipo precedente. Trata-se aqui de uma variável numérica positiva (ou negativa) cujos valores estão contidos num intervalo conhecido. Por exemplo, a variável “altura” ou “peso” em um conjunto de indivíduos é uma variável-intervalo que a teoria implicativa trata, respondendo a duas interrogações do usuário:

- Qual é a melhor subdivisão do intervalo em p sub-intervalos maximizando um critério informacional?

- Entre duas variáveis-intervalo, transformadas em dois conjuntos de p sub-intervalos, quais implicações se pode evidenciar reunindo, se for necessário, alguns destes p sub-intervalos?

Os valores figurando na tabela .csv são numéricos: são os valores assumidos pela variável-intervalo. O software faz, ele mesmo, a partição ótimo do conjunto dos valores segundo uma sub-divisão em um número de sub-intervalos escolhidos pelo usuário.

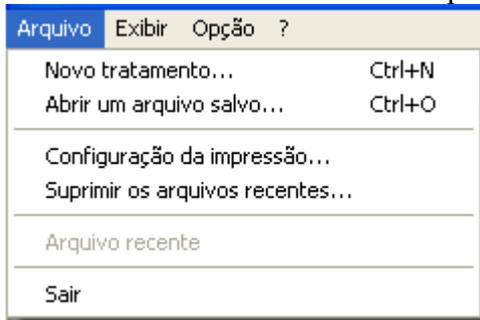
As variáveis-intervalos devem dispor de uma codificação literal ou numérica seguida de um espaço e de um “i”. Assim, uma variável *Idade* de tipo intervalo será, por exemplo, codificada por: *Idade i* em Excel.

- **Variáveis principais e secundárias ou suplementares:** qualificamos uma variável de principal toda variável considerada normalmente pelo CHIC. Como padrão, todas as variáveis são principais. As variáveis suplementares (em geral são as descritivas), são variáveis que não interferem no cálculo das contribuições das categorias. Por exemplo, se desejamos saber se uma implicação é na maior parte formada por pessoas do sexo feminino ou masculino, definimos para cada indivíduo a variável “sexo”. Em seguida, estas variáveis são consideradas apenas quando da procura da contribuição ou da tipicidade das categorias. Para definir uma variável secundária ou suplementar, acrescentamos ao nome da variável um espaço e um “s”. Assim, o atributo *Feminino* é, por exemplo, codificada *Fem* como variável principal e *Fem s* como variável suplementar.

No momento, deve-se fazer esta transformação “manualmente” fora do CHIC, mas está previsto poder fazê-lo com CHIC (em um futuro próximo).

Efetuar um tratamento

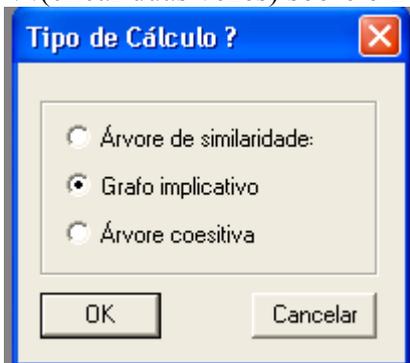
Para efetuar um tratamento: √ Arquivo →



Várias funções estão disponíveis:

√ *Novo tratamento*

√√(clique duas vezes) sobre o nome do arquivo escolhido



Os tratamentos propostos são:

- *Similaridade*: efetua a análise das proximidades segundo I. C. LERMAN, e produz uma janela de resultados numéricos (índices, ...) e uma janela apresentando a árvore hierárquica de similaridades.
- *Grafo implicativo*: efetua os cálculos dos índices de implicação no sentido da análise implicativa, clássica ou entrópica, segundo a opção escolhida, em seguida apresenta uma janela de resultados numéricos (ocorrências, desvio-padrão, coeficientes de correlação) e, em cima, uma janela apresentando um grafo. Os resultados numéricos aparecerão igualmente com os outros tratamentos.
- *Árvore coesiva*: efetua os cálculos dos índices de coesão implicativa no sentido da análise implicativa, depois apresenta uma janela de resultados numéricos e uma janela apresentando uma árvore ascendente segundo o índice decrescente das coesões.

√ *Abrir um arquivo salvo* →

Permite aplicar um modelo ou estado (parâmetros visuais tais que a seleção e a posição dos itens) salvo anteriormente sob o grafo implicativo em uso. O interesse de aplicar um modelo sobre um cálculo está no fato de podermos salvar vários modelos ou estados para o mesmo arquivo. Assim, se desejamos passar de uma representação a outra de um mesmo grafo, não somos obrigados a relançar todos os cálculos.

Para apagar os arquivos da lista de arquivos recentes:

√ *Apaga os arquivos recentes* →



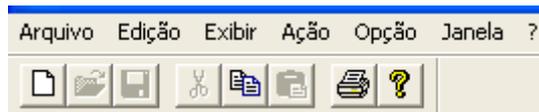
Selecionar um ou vários arquivos e depois clicar sobre *Suprimir*.

Para todos os cálculos, os elementos das partições obtidas pelas variáveis-intervalo são reunidos por adjacências sucessivas. Se por exemplo, uma variável P (como peso) é organizada segundo uma partição de 4 intervalos, P1, P2, P3 e P4, e uma variável T (como altura) é organizada segundo uma partição de 3 intervalos, T1, T2 e T3, o programa tratará, como novas variáveis, todas as reuniões possíveis dos intervalos adjacentes: por exemplo, P23 é obtido pela reunião de P2 e P3; P24 é obtido pela reunião de P2, P3 e P4.

ATENÇÃO !!! Por enquanto, as variáveis-intervalo são limitadas a uma única letra como P e T. CHIC mune estas variáveis de um índice correspondente ao sub-intervalo da variável-intervalo.

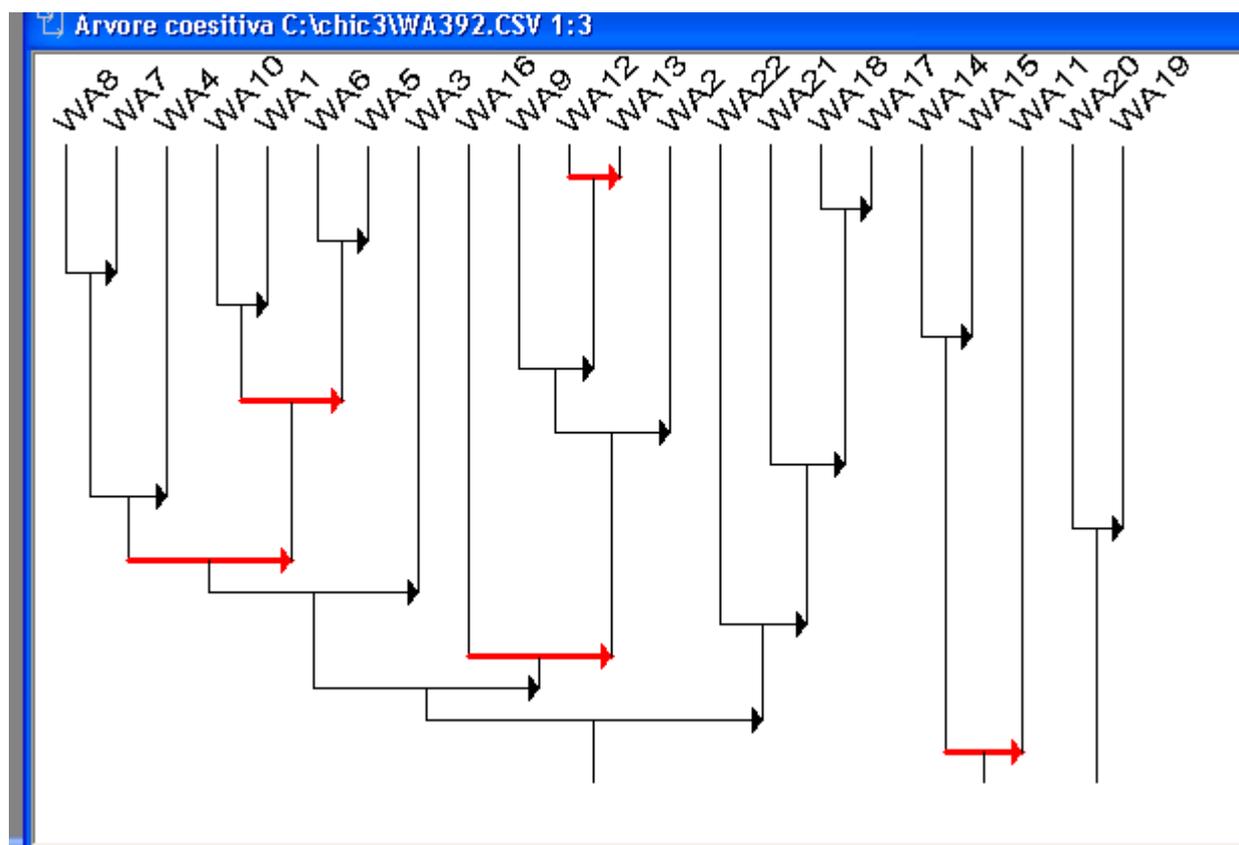
Para passar de um tratamento a outro:

Se for de um arquivo ainda não tratado por CHIC (ou muito anteriormente tratada), ir ao sub-menu *Novo tratamento* do menu *Arquivo* e selecionar como anteriormente o arquivo escolhido. Se for um dos últimos arquivos tratados por CHIC, basta retomá-lo na janela do menu *Arquivo*. No curso dos tratamentos, a barra de menu aparece da seguinte maneira:



Aspectos particulares que dizem respeito aos cálculos e às saídas correspondentes: Quando do lançamento de um dos diferentes tratamentos, a frequência de cada item ou variável é exibida, assim como os coeficientes de correlação linear entre todas as variáveis principais, duas a duas. Em seguida, os índices do tratamento escolhido são exibidos (índice de similaridade, de implicação, ...)

Árvore de similaridades e árvore coesiva.



A árvore aparece em uma janela que dissimula parcialmente ou totalmente os resultados numéricos que figuram em uma outra janela.

Passamos facilmente de uma janela a outra, transitando, por exemplo, pelo menu *Janela*. Os resultados numéricos dos cálculos são exibidos em uma janela do tipo texto e estes podem ser compreendidos unicamente em referência ao tratamento executado.

A figura mostra uma árvore hierárquica com os níveis significativos (flechas mais grossas ou em vermelho). A cada tratamento está associada uma janela de texto que contém os resultados do cálculo inicial.

Obs.: A hierarquia, mesmo que dita ascendente, é representada por um gráfico “descendente”: a base da hierarquia é colocada no alto do gráfico.

Eis a janela de texto:

chic CHIC - [Grafo implicativo C:\chic3\WA392.CSV 1:1]

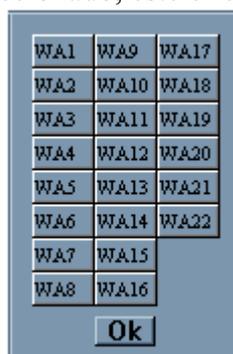
Arquivo Edição Exibir Opção Janela ?

	Occurrence	Média	Desvio padrão
WA1	: 87.00	0.67	0.47
WA2	: 113.00	0.87	0.34
WA3	: 74.00	0.57	0.50
WA4	: 36.00	0.28	0.45
WA5	: 87.00	0.67	0.47
WA6	: 75.00	0.58	0.49
WA7	: 57.00	0.44	0.50
WA8	: 26.00	0.20	0.40
WA9	: 37.00	0.28	0.45
WA10	: 44.00	0.34	0.47
WA11	: 21.00	0.16	0.37
WA12	: 47.00	0.36	0.48
WA13	: 67.00	0.52	0.50
WA14	: 13.00	0.10	0.30
WA15	: 28.00	0.22	0.41
WA16	: 33.00	0.25	0.44
WA17	: 77.00	0.59	0.49
WA18	: 59.00	0.45	0.50
WA19	: 106.00	0.82	0.39
WA20	: 76.00	0.58	0.49
WA21	: 37.00	0.28	0.45
WA22	: 5.00	0.04	0.19

Além disso, uma caixa de ferramentas nomeada barra de itens aparece.

Barra de itens.

Esta caixa apresenta as codificações dos itens tratados. Clicando em um ou mais destas codificações, e clicando em seguida sobre OK, as variáveis correspondentes são deixadas de lado no cálculo, que é retomado para a construção da árvore, desta vez sem estas variáveis. Clicando novamente sobre um item selecionado, este é retomado para um novo cálculo.



Podemos igualmente inverter os itens selecionados, isto é, selecionar os itens não selecionados e reciprocamente:

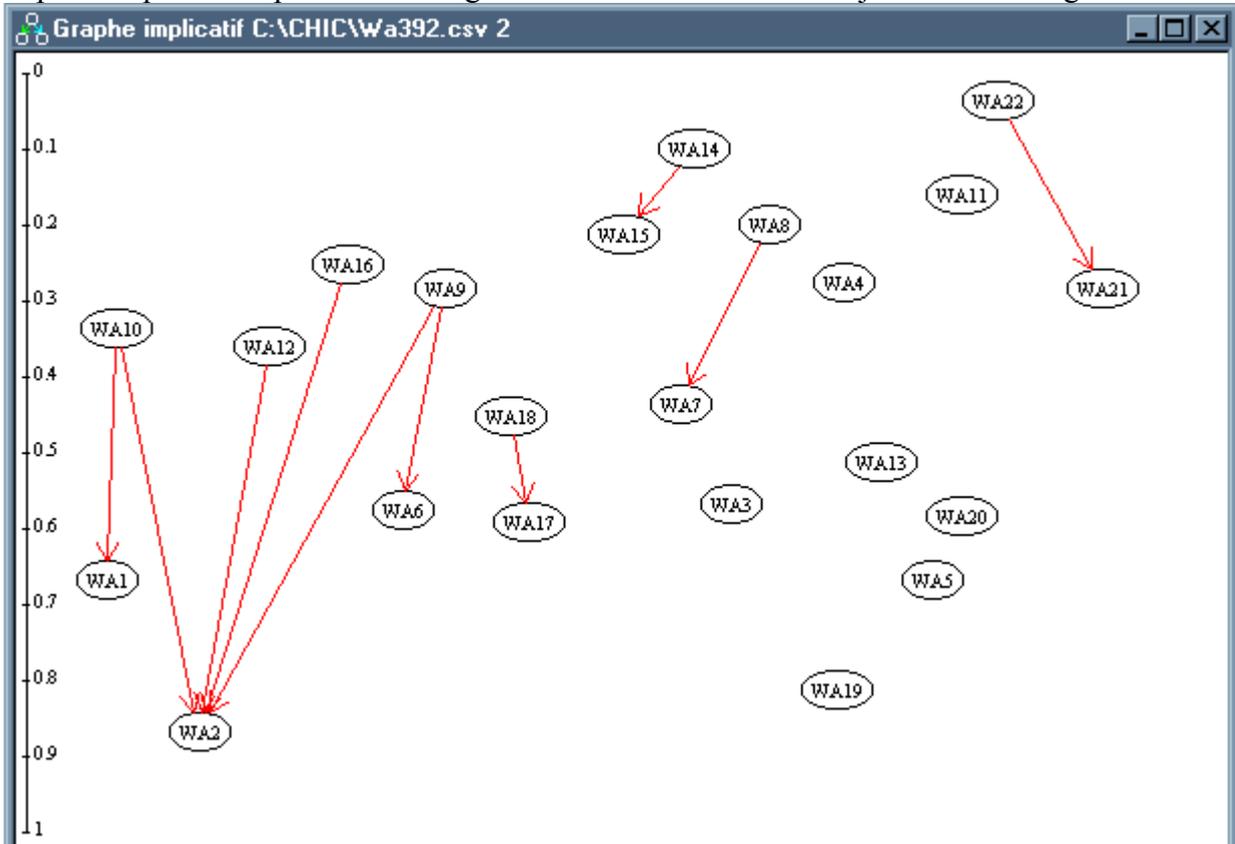
√Ação √Inverter os itens selecionados

Ou Desfazer selecionar todos os itens:

√Ação √Desfazer selecionar todos os itens.

Grafo implicativo

Assim como nas funções similaridade e árvore hierárquica, a função grafo traz uma janela na tela. Essa última contém o grafo implicativo. Atenção, é possível que o grafo esteja vazio. Esta situação se produz quando os parâmetros do grafo não estão corretamente ajustados. Ver logo abaixo.



Somente uma janela é verdadeiramente nova em relação ao que vem sendo escrito acima: é a Barra de Parâmetros...

Barra de parâmetros do grafo implicativo

A janela abaixo aparece, o grafo substitui a janela da árvore anterior. Uma das originalidades desta versão do CHIC é a possibilidade de trabalhar diretamente sobre o grafo e de transformá-lo.



Reorganização do grafo

Basta selecionar um item (clique e manter pressionado) para poder movê-lo. Esse movimento pode somente ser efetuado horizontalmente se escolhermos a opção *respeito das ocorrências* (opção não ativada no momento); ele pode ser efetuado em qualquer direção se escolhermos não respeitar as ocorrências.

O item se movimenta então, mas **TODAS AS LIGAÇÕES SÃO CONSERVADAS** (mesmo as quais não aparecem na tela!). Também é possível selecionar, um a um, um conjunto de itens, ou ainda clicar na janela, movimentar o mouse mantendo o botão pressionado, para selecionar um grupo que será possível de movimentar globalmente.

Esta nova versão do CHIC possui agora um algoritmo que permite desenhar o grafo implicativo de maneira automática. O resultado é interessante em alguns casos, porque obtemos grafo mais rapidamente que movimentando os itens manualmente. Entretanto, em algumas situações, é preferível modificá-lo. É porque o algoritmo é chamado automaticamente e unicamente na criação do grafo com um índice 99. Isto explica que em certos casos, se nenhuma implicação é superior a 99, o grafo é vazio. Uma outra novidade devido à utilização deste algoritmo é que agora, é possível usar curvas de Bézier para representar as implicações. Uma curva de Bézier permite traçar uma curva entre dois itens em vez de uma reta. Ela possui pontos de controle que permitem agir sobre sua forma. Os pontos de controle são visíveis unicamente quando um dos dois itens que compõem a implicação é selecionado. Assim para modificar uma curva, é preciso selecionar um dos dois itens das duas extremidades da implicação, os pontos de controle aparecendo, é possível movimentá-los como movimentamos os itens.

Como padrão, quando é feita a movimentação de um item, CHIC traça uma reta entre o item movimentado e todos os itens a ele ligados. Então os pontos de controle de diferentes curvas seguem as retas. Se desejarmos conservar os pontos de controle nos seus lugares originais, é preciso clicar no botão “c” e o manter pressionado enquanto movimentamos um ou vários itens.

Quando modificamos os índices do grafo implicativo, novas implicações aparecem ou desaparecem. No momento, quando elas aparecem, CHIC coloca os novos itens no alto à esquerda da tela e superpõe os nomes. É o usuário que os movimenta seguindo suas vontades. Se mesmo assim quisermos usar o algoritmo de desenho de grafo automático, chamar:

√ Ação

√ Desenha o grafo de maneira automática

ou pressionar Control + D.

Atenção 1 Se constatarmos que temos implicações demais (muitas para que possamos manipulá-las com o mouse), é preciso saber que para que o algoritmo as localize melhor (seguindo os critérios que ele usa), este pode levar um tempo que varia como uma função exponencial com o número de arcos no grafo. Então, de acordo com a potência do computador usado, este procedimento pode demorar muito tempo (várias horas em certos casos).

Atenção 2 Usando esse algoritmo, o grafo pode ficar grande, ou mesmo muito grande para CHIC. Neste caso, uma janela de diálogo aparecerá pedindo a redução do tamanho do grafo. Por consequência, uma parte do grafo não será representada. Essa situação se produz quando manipulamos grafos muito complexos e, daí parece evidente que não tiraremos uma informação

pertinente desse grafo, esta é a razão do porque suprimir variáveis ou usar índices de implicação maiores.

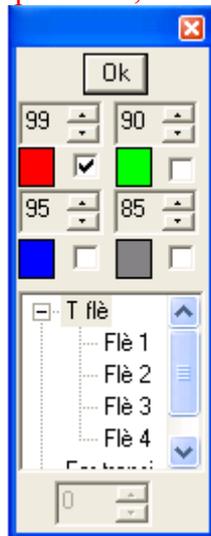
Atenção 3 Quando usamos o algoritmo de desenho de grafo automático, CHIC nem sempre conserva a localização inicial dos itens. É a razão pela qual que este procedimento só aparecerá ao pedido do usuário.

Barra de parâmetros do grafo implicativo

Quando fazemos um primeiro tratamento, esta barra aparece ao mesmo tempo que o grafo. Senão, ir a *Exibir* e clicar no menu correspondente.

O esquema abaixo ilustra o funcionamento da barra de parâmetros do grafo.

(mettre la copie d'écran correspondente, la version portugaise semble incomplete)



Atenção: Após ter modificado as opções da caixa de parâmetros, sempre confirmar clicando em *OK*. Quatro bandas de desfilamento permitem associar cores diferentes a índices diferentes. Se, por exemplo, um primeiro índice é regulado a 99 e é associado à cor vermelha, é possível modificar o índice e a cor. Simultaneamente, podemos selecionar um segundo índice (linha abaixo da primeira seleção), por exemplo 95 e associá-lo a uma cor diferente da anterior.

As opções:

- *T flè*: ela permite modificar o tamanho das flechas normais (não pontilhadas). É possível modificar o tamanho das flechas dos 4 índices, para isso é preciso clicar no pequeno “mais” ao lado de *T flè*, selecionar uma flecha e modificar o tamanho com o elevador.

- *T fer*:

ATENÇÃO OPÇÃO NÃO VALIDA ATUALMENTE

Esta opção permite modificar o tamanho das flechas transitivas (em pontilhado).

- *Fer transi*: Ela permite, escolhendo 1, obter um traçado de todos os arcos associados às implicações transitivas nos índices escolhidos, respeitando as cores. Esses novos arcos aparecem em pontilhado.

- *Com occ.*:

ATENÇÃO OPÇÃO NÃO VALIDA ATUALMENTE

Esta opção permite respeitar ou não as ocorrências no tratamento em curso. Se o valor é 1 os itens podem se mexer somente no sentido horizontal. Se no menu opção, especificamos *respeite as ocorrências do grafo*, um referencial aparece a direita para diferenciar os itens segundo suas ocorrências. Mas assim que pedimos para não respeitar mais as ocorrências do grafo, o referencial desaparece mesmo se recolocarmos a opção *Com occ.*

- *T círculo*:

ATENÇÃO OPÇÃO NÃO VALIDA ATUALMENTE

Ela permite modificar o tamanho dos círculos do grafo.

IMPORTANTE: Os nomes das variáveis que aparecem no grafo na primeira saída de tela são os que introduzimos no arquivo. Esses nomes aparecem na barra de itens.

As pequenas janelas que podem incomodar na leitura da tela podem ser movimentadas a vontade. Ou então elas podem ser, segundo o caso, fechadas ou abertas usando os comandos a seguir:

-F2 para a *barra de itens*

-F3 para a *barra de parâmetros do grafo*

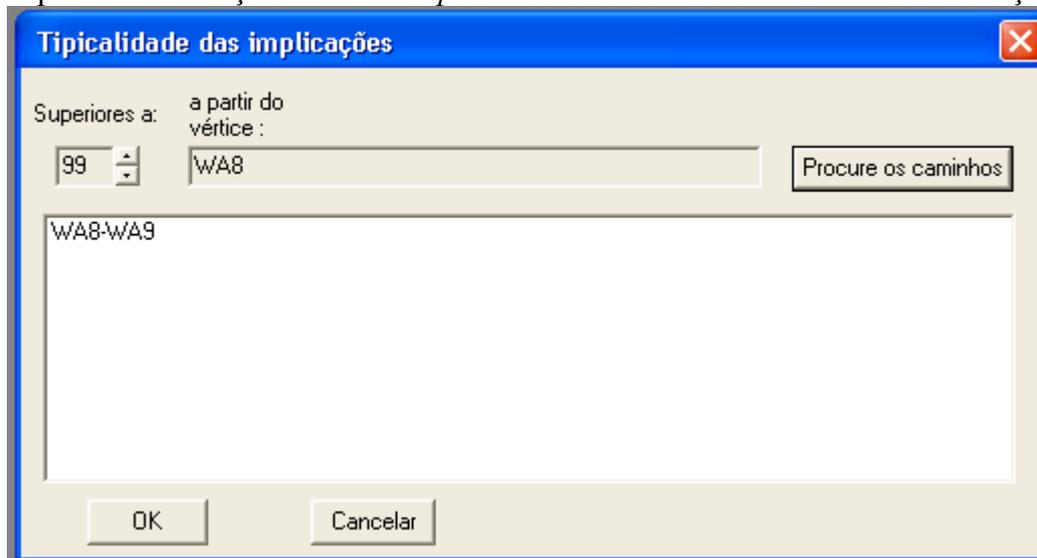
Nos já vimos que, para que isto seja feito, podemos também ir ao menu *Exibir*. No tratamento “grafo implicativo”, a caixa de ferramentas “*barra de itens*” funciona como no caso das árvores.

Cálculo das tipicalidades / contribuições associadas a um caminho

O cálculo das tipicalidades e das contribuições dos indivíduos é possível em relação aos caminhos do grafo implicativo.

Clicamos sobre **UM** pico considerado como origem do caminho.

depois $\sqrt{\text{Ação}}$ $\sqrt{\text{Tipicalidade do indivíduo}}$ ou $\sqrt{\text{Contribuição dos indivíduos}}$



Escolher um índice, e enfim clicar sobre *procure os caminhos*. Os caminhos conexos, de origem o pico escolhido e o índice retido, aparecem então na janela. Podemos selecionar um ou mais, que se sigam ou não (técnica Windows), e depois clicar em OK. Os resultados aparecem então na janela de texto associada ao cálculo.

Obtemos assim a contribuição ou a tipicalidade de cada indivíduo para o(s) caminho(s) escolhido(s), e depois a lista de riscos associados a cada variável suplementar (um risco fraco é o índice de uma forte contribuição ou uma forte tipicalidade) e enfim a variável que contribui mais ou a mais típica.

Grafo implicativo C:\chic3\WA392.CSV 2:1

Tipicalidade ao caminho : WA8-WA9

Tipicalidade dos indivíduos:

e1 : 0.007	e2 : 0.007	e3 : 0.007	e4 : 0.014	e5 : 1.000	e6 :
e7 : 0.007	e8 : 0.007	e9 : 0.014	e10 : 1.000	e11 : 0.007	e12 :
e13 : 0.007	e14 : 1.000	e15 : 0.007	e16 : 0.007	e17 : 0.007	e18 :
e19 : 1.000	e20 : 1.000	e21 : 0.007	e22 : 0.007	e23 : 1.000	e24 :
e25 : 0.007	e26 : 0.014	e27 : 0.007	e28 : 0.007	e29 : 0.014	e30 :
e31 : 0.007	e32 : 0.007	e33 : 0.014	e34 : 0.007	e35 : 0.007	e36 :
e37 : 0.007	e38 : 0.007	e39 : 0.014	e40 : 0.007	e41 : 0.007	e42 :
e43 : 0.007	e44 : 0.007	e45 : 1.000	e46 : 0.014	e47 : 0.007	e48 :
e49 : 1.000	e50 : 0.007	e51 : 1.000	e52 : 0.007	e53 : 0.007	e54 :
e55 : 0.007	e56 : 1.000	e57 : 0.007	e58 : 0.023	e59 : 0.007	e60 :
e61 : 0.007	e62 : 0.007	e63 : 0.007	e64 : 0.007	e65 : 0.007	e66 :
e67 : 0.007	e68 : 0.014	e69 : 0.007	e70 : 0.007	e71 : 0.007	e72 :
e73 : 0.007	e74 : 0.014	e75 : 0.014	e76 : 0.007	e77 : 0.007	e78 :

Cálculo das tipicalidades / contribuições associadas a uma classe

Para obter esta informação quando a árvore aparece, basta pedir a *contribuição* ou a *tipicalidade* na opção *Ação*. Se escolhermos uma, ela vem acompanhada de nomes de indivíduos com o grau de contribuição ou de tipicalidade, e do grupo de indivíduos que mais contribuiu a esta classe ou do grupo mais típico desta classe (grupo ótimo).

Cálculo das tipicalidades / contribuições das categorias de um grafo coesivo

Para obter esta informação quando a árvore aparece, mesmo depois do cálculo precedente, basta pedir na opção "*Ação*" o *cálculo das contribuições das variáveis suplementares* ou o *cálculo das tipicalidades das variáveis suplementares*. As classes aparecem na ordem de suas constituições, acompanhadas das contribuições ou tipicalidades respectivas de cada uma das variáveis, e depois o grupo de variáveis que contribui mais ou o mais típico com seu risco estatístico (repetimos: mais o risco é fraco, mais a confiança é forte).

Complementos

Podemos SALVAR um grafo implicativo, ou IMPRIMIR uma janela (resultados numéricos ou gráficos):

√ *Arquivo* √ *salvar como...* → uma barra de dialogo de salvar se abrirá propondo uma extensão .gra (como grafo) indispensável para que CHIC reconheça o arquivo. Neste caso, CHIC salva a posição dos itens e permite obter o último estado do grafo que tínhamos salvado. Não serão as janelas estatísticas que serão abertas quando abriremos o arquivo usando a função *Abrir um arquivo salvo...* (tomando o cuidado de pedir os arquivos tendo uma extensão .gra)

Podemos também *Copiar um grafo de CHIC* para inseri-lo num aplicativo:

√ *Editar* √ *Copiar*

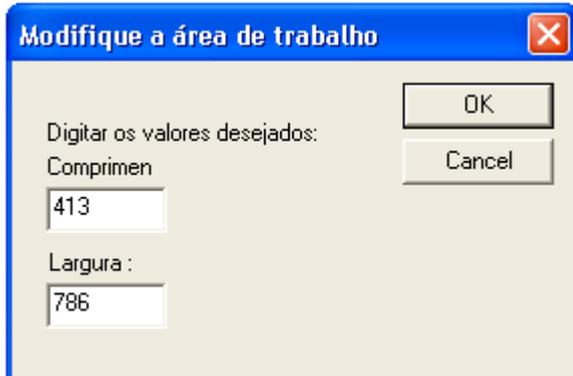
Em seguida podemos colar o grafo em um aplicativo procedendo assim no aplicativo:

√ *Editar* √ *Colar*

Podemos INTERROMPER um cálculo a qualquer momento (ou quase) pressionando o botão “Esc”

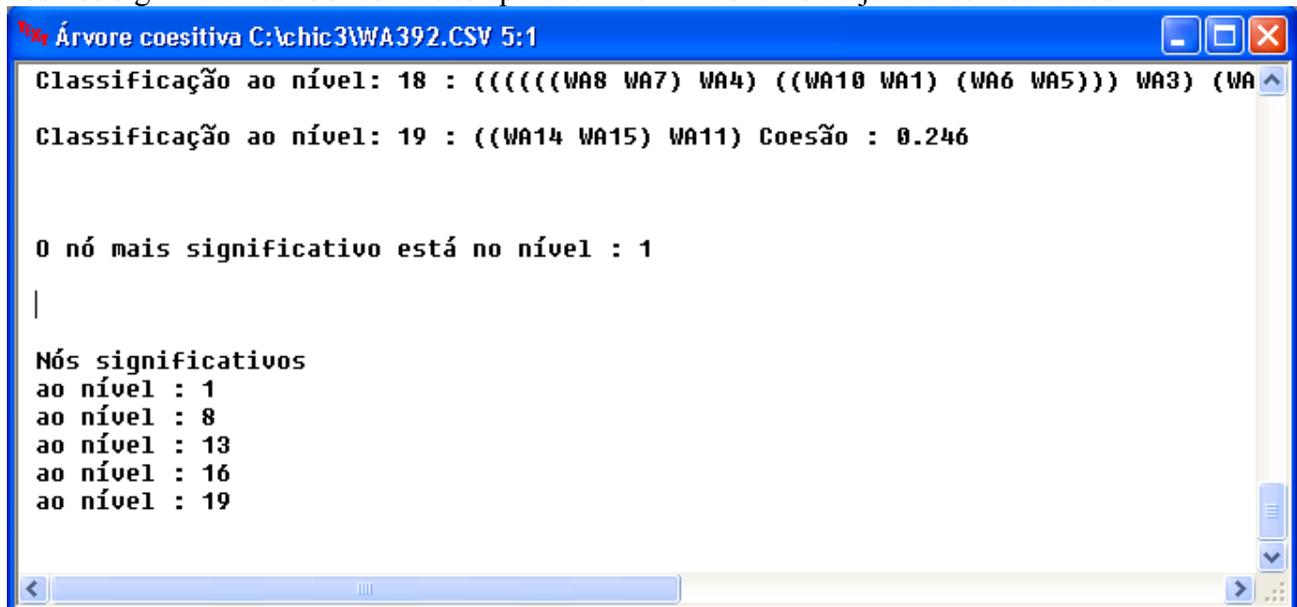
É possível mudar o tamanho da superfície sobre a qual as variáveis do grafo ou da árvore estão localizadas.

√Ação √Modificar superfície de trabalho... →



Nós significativos

Se escolhermos esta opção na caixa de dialogo *Opção*, durante os cálculos relativos à construção do grafo de similaridades e do grafo coesivo, CHIC efetua os cálculos necessários para a determinação dos nós significativos. Os resultados aparecem no fim do texto da janela de resultados.



e na janela do grafo onde cada nível significativo aparece com uma flecha mais grossa e vermelha. A seguir...

Referencias

- Ag Almouloud , S., (1992) : L'ordinateur: outil d'aide à l'apprentissage de la démonstration et de traitement d'analyse de données didactiques. Thèse de l'Université de Rennes 1.
- Bailleul M. (1994) : Analyse statistique implicative : application à la modélisation de l'enseignant dans le système didactique. Thèse, Université de Rennes 1.
- Bodin, A (1996) : Modèles sous-jacents à l'analyse implicative et outils complémentaires,

cahiers du séminaire de didactique de l'IRMAR de Rennes.

- Bodin, A., Couturier, R. et Gras, R (1996) : 'Analyse d'une épreuve de concours par la méthode implicative'. Communication aux journées de la Société Française de Classification, Vannes
- Bodin, A. (1996), 'Improving the Diagnostic and Didactic Meaningfulness of Mathematics Assessment in France' Annual Meeting of the American Educational Research Association AERA - New-York
- Bodin, A., Gras, R. et Lagrange, J.B. (1997) : Implication statistique, Prépublication IRMAR n° 97-32, Rennes
- Couturier, R. et Gras R. (1999) : Introduction de variables supplémentaires dans une hiérarchie de classes et application à CHIC, Actes des 7èmes Rencontres de la Société Francophone de Classification, 87-92, Nancy, 15-17 septembre 1999
- Gras, R et al (1996) : L'implication statistique. Nouvelle méthode exploratoire de données. La Pensée Sauvage. Grenoble
- Gras, R et Pécal, M. (1995) : L'évaluation en mathématiques : perspectives institutionnelles, pédagogiques et statistiques. Actes de l'université d'été de l'APMEP - Sophia Antipolis 10-14 juillet 1995 - Brochure N° 102 de l'APMEP.
- Gras, R. (1986) : Recherches sur l'apprentissage : Analyse des correspondances et méthodes statistiques apparentées - Cahier du Cirade - Université du Québec à Montréal.
- Gras, R. (1992) : Data analysis : a method for the processing of didactic questions. In Research in Didactique of mathematics - selected papers - Douady, R. & Mercier, A. Ed - La Pensée Sauvage – Grenoble
- Gras, R. (1995), Méthodes d'analyses statistiques multidimensionnelles en didactique des mathématiques. Actes du colloque ARDM de Caen (27 - 29 janvier 1995) - publié par l'ARDM
- Gras, R. , Larher, A. (1992) : 'L'implication statistique, une nouvelle méthode d'analyse de données', Mathématique, Informatique et Sciences Humaines, n° 120.
- Gras, R.(1992) : L'analyse des données: une méthodologie de traitement de questions de didactique, Recherches en Didactique des Mathématiques, Vol. 12-1.
- Gras, R et Ratsimba-Rajohn (1996) : Analyse non symétrique de données par l'implication statistique, RAIRO, Recherche Opérationnelle, n°3-96, AFCET Paris.
- Gras R., Briand H., Peter P., Philippe J. (1997) : Implicative statistical analysis, Proceedings of International Congress I.F.C.S., 96, Kobe, Springer-Verlag, Tokyo.
- Gras R., Richeton J.P. (2000) : Eléments d'analyse de l'expérimentation d'épreuves de mathématiques en classe de première, Bulletin n° 427 de l'Association des Professeurs de Mathématiques de l'Enseignement Public, 187-201, Paris, ISSN 0240-5709
- Gras R. (2000) : Quelques principes majeurs pour l'élaboration d'un programme de mathématiques pour le second cycle, Bulletin n° 429 de l'Association des Professeurs de Mathématiques de l'Enseignement Public, 522-527, Paris, ISSN 0240-5709
- Gras R., Kuntz P., Couturier R. et Guillet F. (2001) : Une version entropique de l'intensité d'implication pour les corpus volumineux, Proceedings des Journées E.C.D. de Nantes (2001), Hermès
- Gras Robin, Gras Régis, et al : Classification automatique de protéines par un algorithme génétique pour l'optimisation d'un outil d'identification de protéines par empreinte de masses peptidiques, soumis à Mathématiques et Sciences Humaines
- Gras R., Kuntz P. et Briand H. (2001) : Les fondements de l'analyse statistique implicative, Mathématiques et Sciences Humaines, n° 154-155
- Gras R., Diday E., Kuntz P. et Couturier R. (2001) : Variables sur intervalles et variables-intervalles en analyse implicative, Actes du 8ème Congrès de la SFC de Pointe à Pitre,

17-21 décembre 2001, pp 166-173

- Gras Régis, Guillet F., Gras Robin et Philippé J. (2002) : Réduction des colonnes d'un tableau de données par quasi-équivalence entre variables, Extraction des connaissances et apprentissage, Hermès, Volume 1, n°4/2001, p 197-202, ISBN 2-7462-0406-1
- Larher A. (1991) : Implication statistique et applications à l'analyse de démarches de preuve mathématique, Thèse de l'Université de Rennes 1.
- Lerman I.C., Gras R. et Rostam H., (1981) : Elaboration et évaluation d'un indice d'implication pour des données binaires, I et II, Mathématiques et sciences Humaines, n°75, Paris
- Peter P., Gras R., Philippé J. et Baquédano S. (2001): L'analyse implicative pour l'étude d'un questionnaire de personnalité, Proceedings des Journées E.C.D. de Nantes 2001, Hermès
- Polo, M. (1996) : Le repère cartésien dans les systèmes scolaires français et italien : étude didactique et application de méthodes d'analyse statistique multidimensionnelle, Thèse de l'Université de Rennes 1
- Ratsimba-Rajohn, H. (1992) : Contribution à l'étude de la hiérarchie implicative, application à l'analyse de la gestion didactique des phénomènes d'ostension et de contradiction, Thèse de l'Université de Rennes 1
- Totohasina, A. (1992) : Méthode implicative en analyse de données et application à l'analyse de conceptions d'étudiants sur la notion de probabilité conditionnelle, Thèse de l'Université de Rennes 1

Apêndice

Métodos de análise de dados praticados no software CHIC

Estas notas, intuitivas e poucas técnicas, têm por finalidade guiar o usuário em suas primeiras interpretações. Os argumentos teóricos são apresentados nas obras e artigos citados nas referências acima.

Análise das similaridades segundo I.C. Lerman

Indícios de similaridade

Como em todos os métodos de classificação, procuramos constituir, em um conjunto V das variáveis, partições de V cada vez menos finas, construídas de maneira ascendente. Essas partições encaixadas são representadas por uma árvore construída usando um critério de similaridade ou de semelhança estatística entre variáveis. A similaridade se define a partir do cruzamento do conjunto V das variáveis com um conjunto E de sujeitos (ou de objetos). Este tipo de análise permite ao usuário estudar e depois interpretar, em termos de tipologia e de semelhança (e não semelhança) decrescente, classes de variáveis, constituídas significativamente a certos níveis da árvore e se opondo a outros nestes mesmos níveis.

O **critério de similaridade** se exprime da maneira seguinte nos casos das variáveis binárias (presença – ausência, verdadeiro – falso, sim – não, etc...):

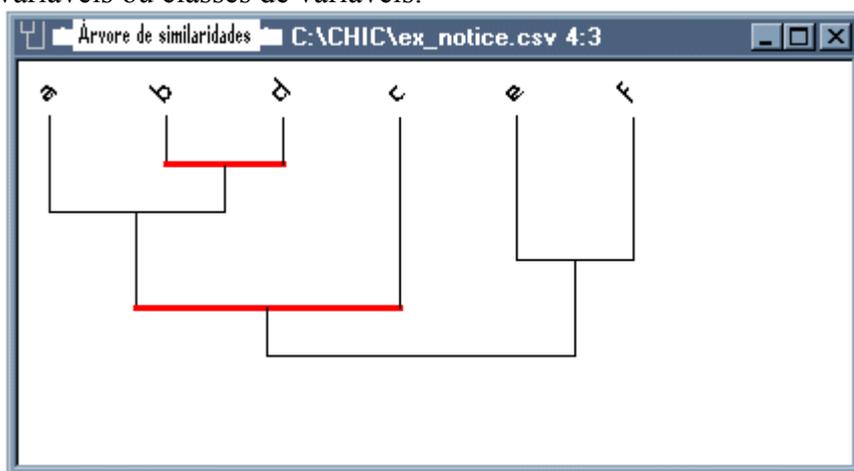
2 variáveis a e b , satisfeitas respectivamente por sub-conjuntos (suportes) A e B de E , são muito semelhantes quando o número k dos sujeitos que os verificam simultaneamente (ou seja os

elementos de $A \wedge B$) é importante de um lado, pelo que teria sido no caso da ausência de ligação entre \underline{a} e \underline{b} , e por outro lado, com relação aos cardinais de E, A e B. Medimos esta semelhança pela probabilidade que k seja superior ao número aleatório esperado nesta situação na qual somente o acaso interviria. O índice correspondente entre as variáveis não é então modificado, desviado pelo tamanho de $A \wedge B$ e não coincide então com o coeficiente de correlação linear.

A modelagem probabilista da variável aleatória, cujo k é a realização presente, pode ser **binomial** ou de Poisson à escolha do usuário. A segunda supõe que E seja uma amostra de uma população mãe mais ampla, o que a primeira não supõe. Se E não tem nenhuma razão estatística a priori de ser representativo, é preferível usar o modelo binomial que analisa a estrutura de E enquanto tal. Quando os parâmetros o permitem, uma aproximação gaussiana destas duas leis é efetuada. O índice de similaridade entre variáveis serve em seguida para definir um índice de similaridade entre duas classes de variáveis segundo este mesmo princípio de comparação entre a observação e o que seria dado pelo acaso. Um índice, dito de coesão, permite não mais reagrupar as classes quando esse reagrupamento é feito “contra – natureza”, isto é, quando o índice de similaridade entre as classes, em processo de reagrupamento, apresenta um índice de coesão muito fraco.

Árvore de similaridade

Assim, para construir uma árvore de similaridade, reunimos em uma classe de primeiro nível, primeiramente, as 2 variáveis que são mais similares no sentido do índice de similaridade, depois 2 outras variáveis ou uma variável e a classe já formada no sentido do índice da classe, e depois outras variáveis ou classes de variáveis.



Duas outras informações são susceptíveis de ajudar na interpretação da árvore: a typicalidade e a contribuição. Falaremos do assunto um pouco mais adiante com a teoria implicativa.

Typicalidade

Certos sujeitos são típicos do comportamento do conjunto da população no sentido seguinte: no estudo da similaridade, eles atribuem ao conjunto das variáveis valores compatíveis com as similaridades constituídas sobre essas variáveis pela população. Se as variáveis suplementares foram definidas pelo usuário, obteremos assim a typicalidade dessas variáveis a partir das typicalidades dos indivíduos que as satisfazem. Por exemplo, no decorrer de uma pesquisa sócio-profissional, serão os auxiliares administrativos que serão típicos do comportamento de uma população de pessoas ativas.

Contribuição

É possível conhecer a contribuição a cada uma das classes de cada um dos sujeitos e então das variáveis suplementares. Cada uma dessas últimas contribui mais ou menos na formação da classe: isto significa que os valores que eles dão às variáveis vão no sentido de suas similaridades. R. Gras e H. Ratsimba-Rajohn elaboraram um critério que permite avaliar essa contribuição relativamente a cada uma das classes. Cada uma delas contribui. Por exemplo, em um questionário de atitude, podemos evidenciar, a “responsabilidade” das mulheres de idade entre 30 e 40 anos na existência de uma certa classe de variáveis principais (ou ativas), o que quer dizer que elas participaram na construção da hierarquia.

Análise das implicações entre variáveis e classe de variáveis

Índices de implicação

O estudo continua sendo feito sobre o cruzamento de um conjunto de variáveis V e de um conjunto de sujeitos E . No caso prototípico das variáveis binárias, queremos dar um sentido estatístico a expressões como: “quando se observa sobre um sujeito de E a variável \underline{a} , em geral observa-se a variável \underline{b} ”. Trata-se então de procurar um modelo estatístico de uma quase implicação do tipo: “Se \underline{a} então quase \underline{b} ”, a implicação lógica estrita sendo raramente satisfeita. A esta quase implicação é associada semanticamente uma **regra**, um tipo de teorema que liga uma premissa e uma conclusão. Vemos assim a diferença entre o método de análise de similaridades que é simétrico e o método implicativo que é, por essência, não simétrico.

Partindo dos sub-conjuntos A e B , suportes respectivos de \underline{a} e \underline{b} , nós interessamos na medida do sub-conjunto dos contra-exemplos da implicação, a saber as ocorrências da propriedade $(a \wedge \neg b)$ do suporte $A \wedge \neg B$ ($\neg B$ sendo o complementar de B em E). O número k de contra-exemplos é considerado como a realização de uma variável aleatória de um modelo de **Poisson** ou de um modelo **binomial**, um e outro aproximados pela lei de Gauss quando é legitimado pelos parâmetros. O modelo de Poisson é mais severo que o modelo binomial.

Intuitivamente, diremos que a implicação é admissível no índice de confiança α se a probabilidade que essa variável aleatória seja superior a k é ela mesma superior a $1-\alpha$. Isto é, quanto mais k for pequeno, em relação as ocorrências de \underline{a} e \underline{b} e o tamanho de E , mais a implicação é surpreendentemente grande, então admissível e, sem dúvidas, portadora de um sentido. O número $1-\alpha$ é o **índice de implicação** dito **da teoria clássica**. O valor 0.95 representa um bom valor de admissibilidade quando n , \underline{a} e \underline{b} ultrapassam muitas dezenas de unidades.

Portanto, quando o tamanho das amostras alcança várias centenas, ver milhares ou mesmo centenas de milhares, dispomos de uma modelagem mais complexa, mas mais adequada, pois ela permite estimar não somente a qualidade da implicação direta de $a \Rightarrow b$, mas igualmente sua recíproca $\neg b \Rightarrow \neg a$. Esta modelagem é chamada entrópica, pois ela faz apelo a qualidade da informação recolhida pelos desequilíbrios respectivos dos casos $((a \wedge b)$ e $(a \wedge \neg b)$, e depois $((\neg a \wedge \neg b)$ e $(a \wedge \neg b)$, desequilíbrio que mede a **entropia** no sentido de Shannon. O índice que o corresponde é chamado de **índice de implicação - inclusão** pois ele mede mais fielmente a quase inclusão de A em B.

Grafo implicativo

Um grafo implicativo traduz graficamente a rede de relações quase implicativas entre as variáveis de V. O intervalo de confiança da aparição de arcos ou flechas do grafo é controlável pelo usuário que pode, a sua vontade, aumentar ou diminuir seu número. A transitividade, que pilota a interpretação em termos de caminhos, é aceita a um intervalo de confiança de 0,50.

Durante a análise, podemos nos concentrar unicamente na procura de arcos em “**Amon**” (“pais” ou fontes) de um pico de um grafo e em “**aval**” (“filhos” ou “crianças”) deste mesmo pico. Para isto, basta pedir um cone de origem o pico escolhido. A partir da opção de menu, mas igualmente durante este trabalho, é possível mudar o tamanho da janela de trabalho, o que permite se concentrar na organização dos arcos sobre uma parte do grafo. Além disso, o software sendo bem conhecido “APRIORI”, CHIC permite estudar as conjunções das variáveis. Para isto, procuraremos entre as conjunções de 2 variáveis (então 3 variáveis em jogo: conjunção de 2 variáveis implicando uma variável), 3,4, etc. (respectivamente 4, 5, etc. variáveis em jogo), as que apresentam uma originalidade dada. Esse índice leva em consideração a implicação, implicação entrópica, o suporte das variáveis e uma certa “confidência”. Por exemplo, se pedimos, a um intervalo de confiança de originalidade de 0.80, considerar todas as conjunções pondo 5 variáveis em jogo, seja a conjunção de 4 para a qual procuramos a implicação com a 5°, CHIC calculará todas as implicações possíveis das conjunções de 2, 3 e 4 variáveis retendo as que aparecem no intervalo de confiança de 0.80. Se este intervalo é mudado, o grafo logicamente também o será.

Árvore

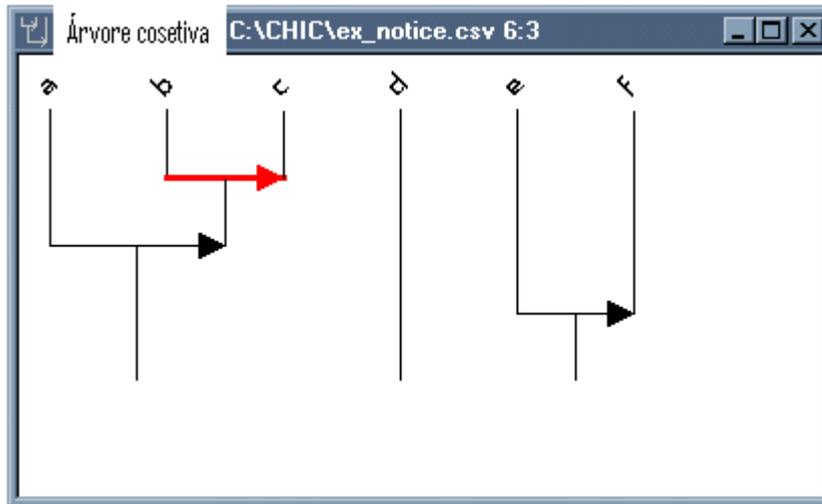
O índice de implicação entre duas variáveis é estendido ao cálculo da coesão da classe. Esta última dá conta da qualidade da implicação orientada dentro de uma classe de variáveis e traduz a noção de meta-regra ou regra sobre regra. Uma hierarquia ascendente ou árvore coesiva traduz graficamente o encaixamento sucessivo das classes constituídas segundo o critério de coesão que é decrescente segundo os níveis (no sentido contrário da formação das classes de variáveis) da hierarquia. Um intervalo de confiança de parada sobre a coesão permite evitar a constituição das classes que não têm sentido implicativo, o que não se produz nas hierarquias clássicas, mas fica mais conforme a semântica.

Níveis e nós significativos

As noções de nível e de nós significativos, como precedentemente, sublinhados por uma flecha vermelha assinala ao usuário as classes sobre os quais ele deve ter mais atenção no fato de sua melhor conformidade com os indícios de implicação iniciais.

Na representação abaixo, observamos que no primeiro nível, se forma uma classe ordenada (b,c) do fato que a implicação de b sobre c é a mais forte entre todas as implicações possíveis entre variáveis. A ele, corresponde necessariamente um nó significativo. Em seguida, no nível 2 uma

meta-regra aparece de a sobre (b,c). Ela se interpreta, por exemplo, da maneira seguinte: se a é verdadeiro então (se b é verdadeiro então c) é geralmente; $(a \Rightarrow b) \Rightarrow c$) é equivalente a $a \wedge b \Rightarrow c$. No nível 4 se forma a regra (e,f). A variável d não implica e não é implicada por nenhuma outra.



Tipicalidade e contribuição

Que se trate dos caminhos do grafo implicativo ou das classes coesivas, é interessante conhecer qual é a responsabilidade dos sujeitos e das variáveis suplementares em suas formações, como foi feito para a similaridade. Esta opção é possível, de duas maneiras:

- primeiramente, pelo cálculo do valor da tipicalidade de um sujeito x caracterizando sua conformidade ou sua quase conformidade à tendência geral dada pela intensidade da implicação
- inclusão de uma variável \underline{a} sobre uma variável b. Por exemplo, se x toma o valor $a(x)=0,2$ segundo \underline{a} e o valor $b(x)=0,9$ segundo \underline{b} , sua responsabilidade com relação à implicação $a \Rightarrow b$ é 0,73. Além de mais, se a intensidade da implicação de \underline{a} sobre \underline{b} é 0,75, x é mais típico que o sujeito y que teria a responsabilidade de 0,95. Definimos alias a distância de x à regra $a \Rightarrow b$ a partir desta responsabilidade. Essa distância varia entre 0 e 1. O valor da tipicalidade é o complemento a 1 desta distância. Ela pode ser estendida ao conjunto de relações de um caminho do grafo implicativo ou ao de uma classe da hierarquia coesiva. Os sujeitos que teriam um valor muito bom de tipicalidade poderiam ser considerados como prototípicos da população. Podemos saber qual é o grupo ótimo dos sujeitos que são os mais típicos de um caminho ou de uma classe e tirar a variável suplementar a mais típica deste caminho ou desta classe.
- Em seguida, pelo cálculo da conformidade lógica de um sujeito x à existência de um arco do grafo levando em consideração o intervalo de confiança escolhido, ou da hierarquia. Por exemplo, se o arco (a, b) aparece sobre o grafo ou na árvore, qual que seja a intensidade da implicação de \underline{a} sobre \underline{b} , diremos que esta conformidade é igual a 1 e que ela é igual a 0 no caso contrário. Daí deduz-se a distância de x e a contribuição de x à regra $a \Rightarrow b$ é igual ao complemento desta distância. Estendida a um caminho e a uma classe, ela permite estabelecer o grupo ótimo contributivo, e depois a variável suplementar a mais contributiva ao caminho ou à classe. Essas informações são úteis para orientar o usuário para analisar a ligação de tal ou tal grupo de sujeitos relativamente às regras ou meta-regras particulares.