Módulo 18- Análise de Cluster

Tutorial SPSS – Preparação dos Dados e Geração de Tabelas

Método Hierárquico e Não-Hierárquico

Situação Problema

Uma varejista de roupas e acessórios femininos voltados para a classe A e B iniciará um programa de relacionamento com os clientes, oferecendo atendimento personalizado, promoções específicas e facilidades para cada grupo de clientes. É, portanto, necessário segmentar a clientela em grupos distintos, agrupando clientes com perfil semelhante, para que se possa direcionar a oferta.

A empresa elaborou um questionário com afirmações sobre compras e selecionou uma amostra, ao acaso, de 20 clientes para respondê-los. As questões são mensuradas através de uma nota de 1 a 5, onde 1 representa a nota mais baixa (discordância total) e a nota 5 representa a maior nota (concordância total), conforme é mostrado a seguir:

1) Só compro roupas quando realmente preciso	1	2	3	4	5
2) Uso as roupas que estão na moda	1	2	3	4	5
3) Compro roupas extravagantes mesmo não vá utilizá-las	1	2	3	4	5
4) Quando compro roupas costumo escolher também outras peças e acessórios para combinar	1	2	3	4	5
5) Costumo comprar mais roupas "curinga", fáceis de combinar	1	2	3	4	5
6) Quando gosto não me importo com o preço da peça	1	2	3	4	5
7) Procuro comprar sempre peças exclusivas	1	2	3	4	5
8) Só compro roupas se elas forem de marcas famosas	1	2	3	4	5

Tabela 1: Instrumento de pesquisa.

Para realizar o agrupamento devemos utilizar a ferramenta análise de clusters, conhecida também como análise de conglomerados. Conforme o texto teórico deste módulo, existi basicamente dois grandes grupos de métodos de clusterização: métodos hierárquicos e

métodos não-hierárquicos. A forma de gerar tabelas, bem como de analisar os resultados possuem algumas diferenças. Assim, este tutorial foi dividido em quatro partes: geração de tabelas para o método hierárquico, análise dos resultados para o método hierárquico, geração de tabelas para o método não- hierárquico, análise dos resultados para o método não- hierárquico.

Nesta parte apresentamos a geração de tabelas para ambos os métodos (hierárquico e não-hierárquico)

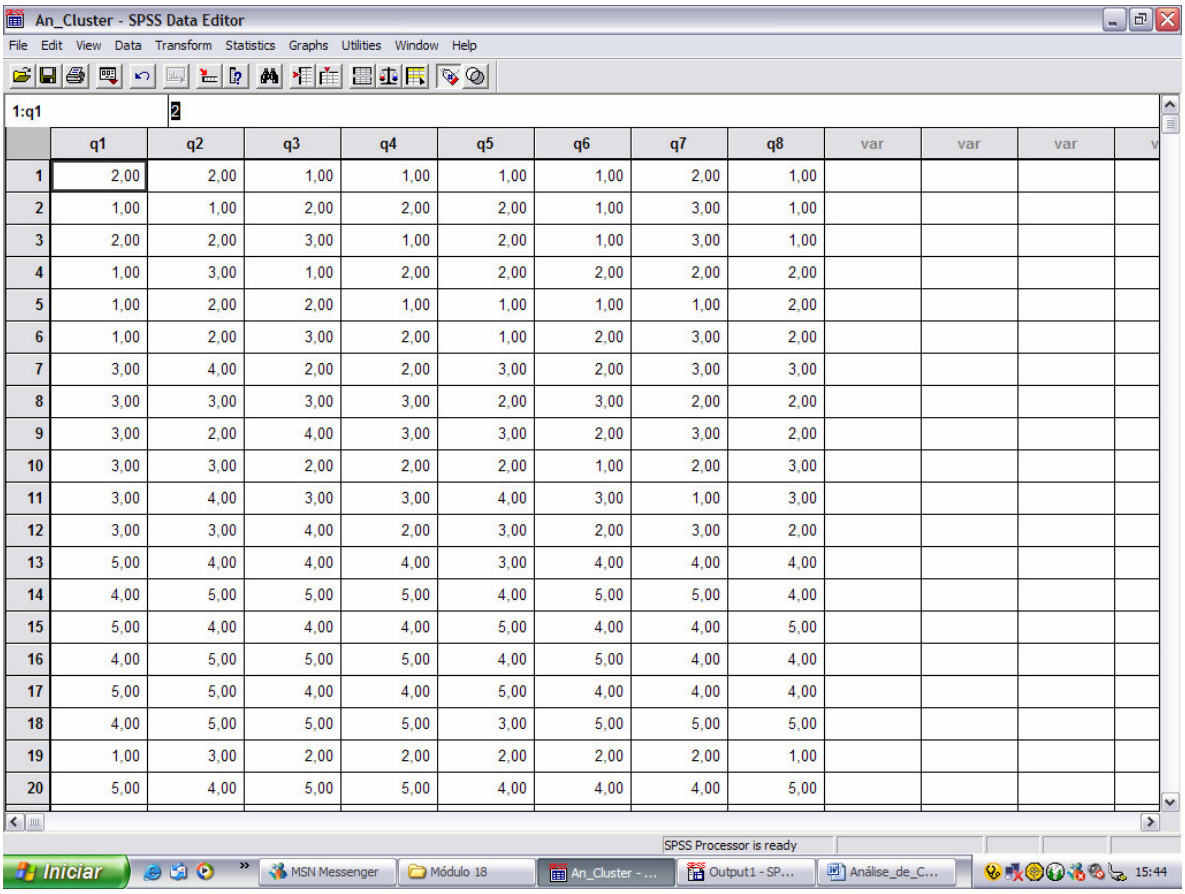
Método Hierárquico

Preparação dos dados

A análise de clusters é uma técnica de interdependência, isto é, não existe uma variável independente ou dependente. Todas as variáveis se relacionam, positivamente ou negativamente, mas nenhuma delas possui relação de dependência com as outras.

Neste problema específico, as afirmações não dependem umas das outras, mas possuem alguma relação. A ferramenta encontrará grupos de mulheres que possuem as mesmas opiniões em relação às afirmações.

Lembre-se que no SPSS as linhas representam os casos (questionários respondidos) e as colunas representam as variáveis medidas (questões do questionário). Os dados devem se inseridos na planilha do SPSS como mostra a figura 1.

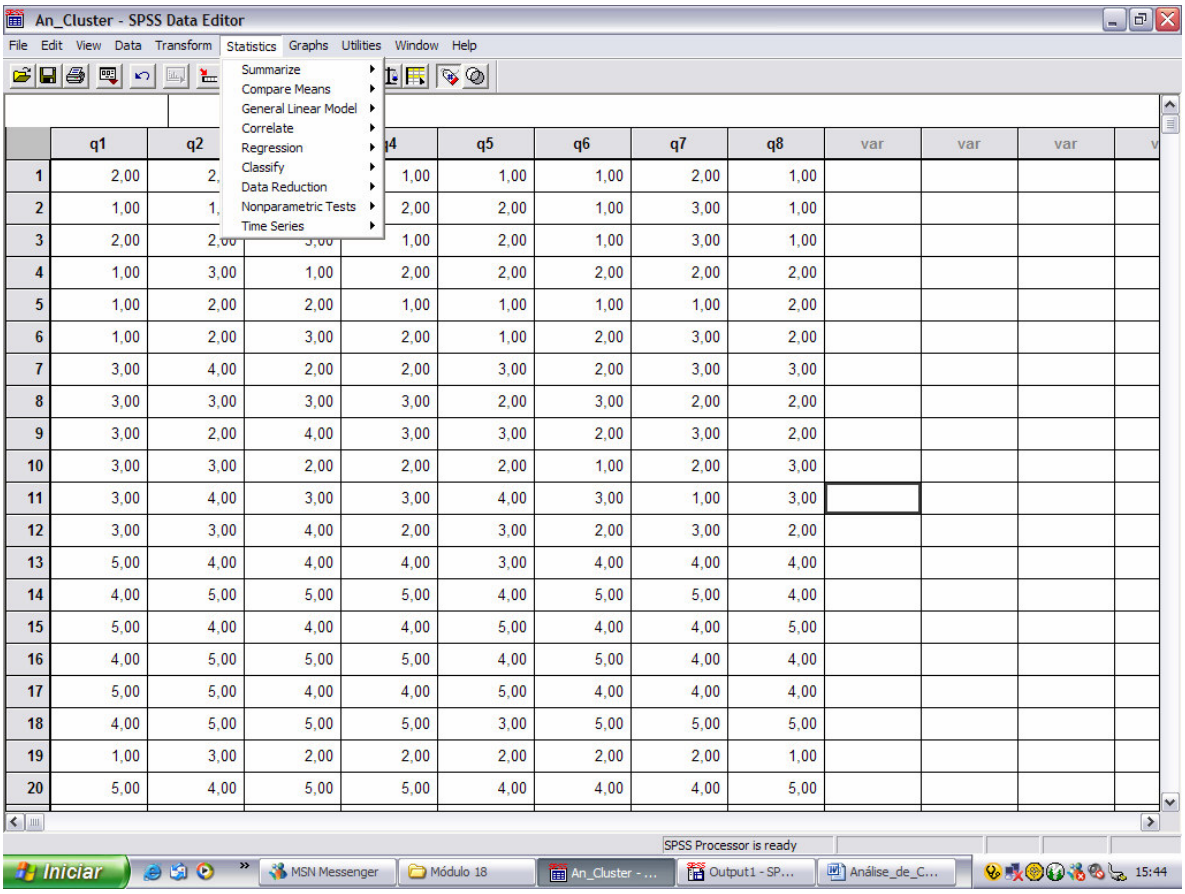


	q1	q2	q3	q4	q5	q6	q7	q8	var	var	var
1	2,00	2,00	1,00	1,00	1,00	1,00	2,00	1,00			
2	1,00	1,00	2,00	2,00	2,00	1,00	3,00	1,00			
3	2,00	2,00	3,00	1,00	2,00	1,00	3,00	1,00			
4	1,00	3,00	1,00	2,00	2,00	2,00	2,00	2,00			
5	1,00	2,00	2,00	1,00	1,00	1,00	1,00	2,00			
6	1,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00			
7	3,00	4,00	2,00	2,00	3,00	2,00	3,00	3,00			
8	3,00	3,00	3,00	3,00	2,00	3,00	2,00	2,00			
9	3,00	2,00	4,00	3,00	3,00	2,00	3,00	2,00			
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00			
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00			
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00			
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00			
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00			
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00			
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00			
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00			
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00			
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00			
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00			

Figura 1: Inserção dos dados na planilha.

Parte 1- Geração de tabelas

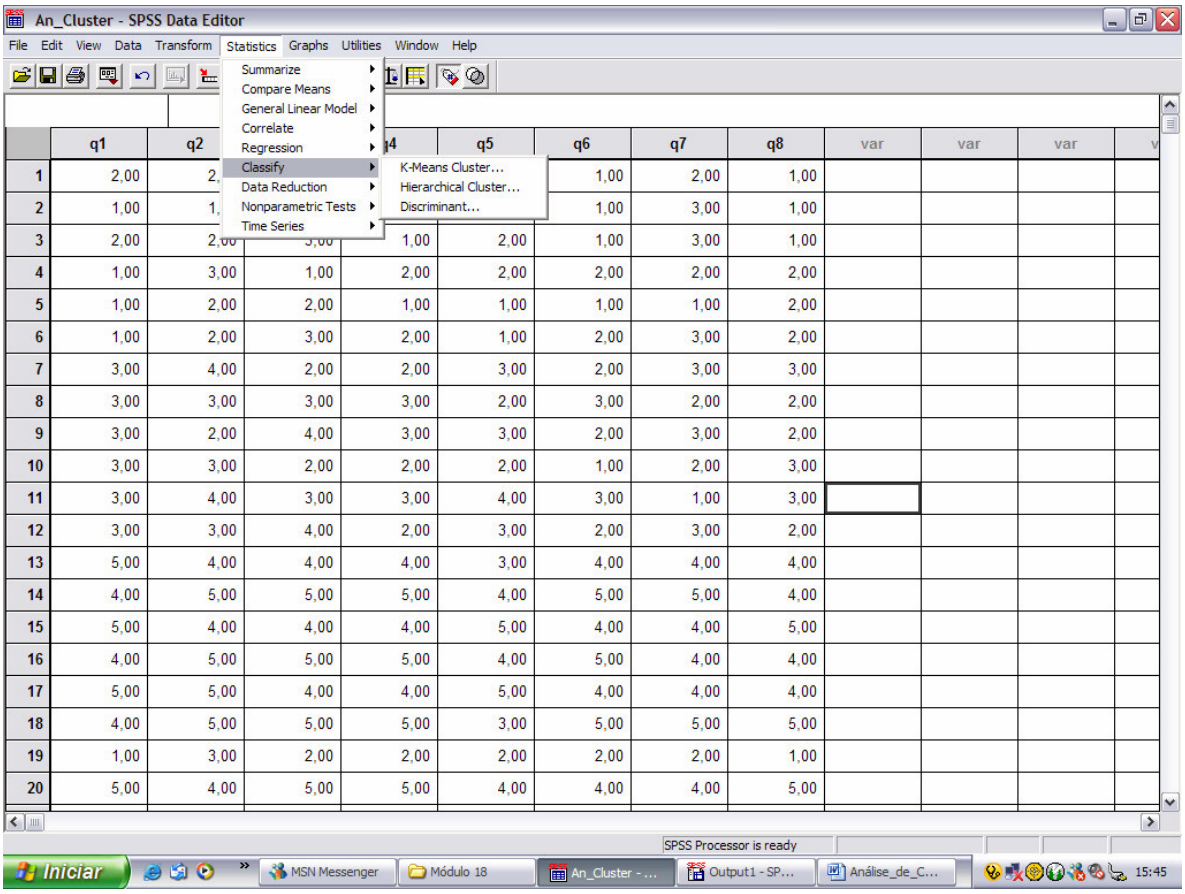
Escolhemos o menu “statistics” na barra de ferramentas.



	q1	q2		q4	q5	q6	q7	q8	var	var	var
1	2,00	2,		1,00	1,00	1,00	2,00	1,00			
2	1,00	1,		2,00	2,00	1,00	3,00	1,00			
3	2,00	2,00		1,00	2,00	1,00	3,00	1,00			
4	1,00	3,00	1,00	2,00	2,00	2,00	2,00	2,00			
5	1,00	2,00	2,00	1,00	1,00	1,00	1,00	2,00			
6	1,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00			
7	3,00	4,00	2,00	2,00	3,00	2,00	3,00	3,00			
8	3,00	3,00	3,00	3,00	2,00	3,00	2,00	2,00			
9	3,00	2,00	4,00	3,00	3,00	2,00	3,00	2,00			
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00			
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00			
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00			
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00			
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00			
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00			
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00			
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00			
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00			
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00			
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00			

Figura 2: Escolha do menu “statistics”.

Como vamos realizar uma análise de cluster, que vai classificar as clientes de acordo com as respostas dadas às afirmações do questionário, devemos escolher a opção “classify”.



	q1	q2			q5	q6	q7	q8	var	var	var
1	2,00	2,				1,00	2,00	1,00			
2	1,00	1,				1,00	3,00	1,00			
3	2,00	2,00		1,00	2,00	1,00	3,00	1,00			
4	1,00	3,00	1,00	2,00	2,00	2,00	2,00	2,00			
5	1,00	2,00	2,00	1,00	1,00	1,00	1,00	2,00			
6	1,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00			
7	3,00	4,00	2,00	2,00	3,00	2,00	3,00	3,00			
8	3,00	3,00	3,00	3,00	2,00	3,00	2,00	2,00			
9	3,00	2,00	4,00	3,00	3,00	2,00	3,00	2,00			
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00			
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00			
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00			
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00			
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00			
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00			
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00			
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00			
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00			
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00			
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00			

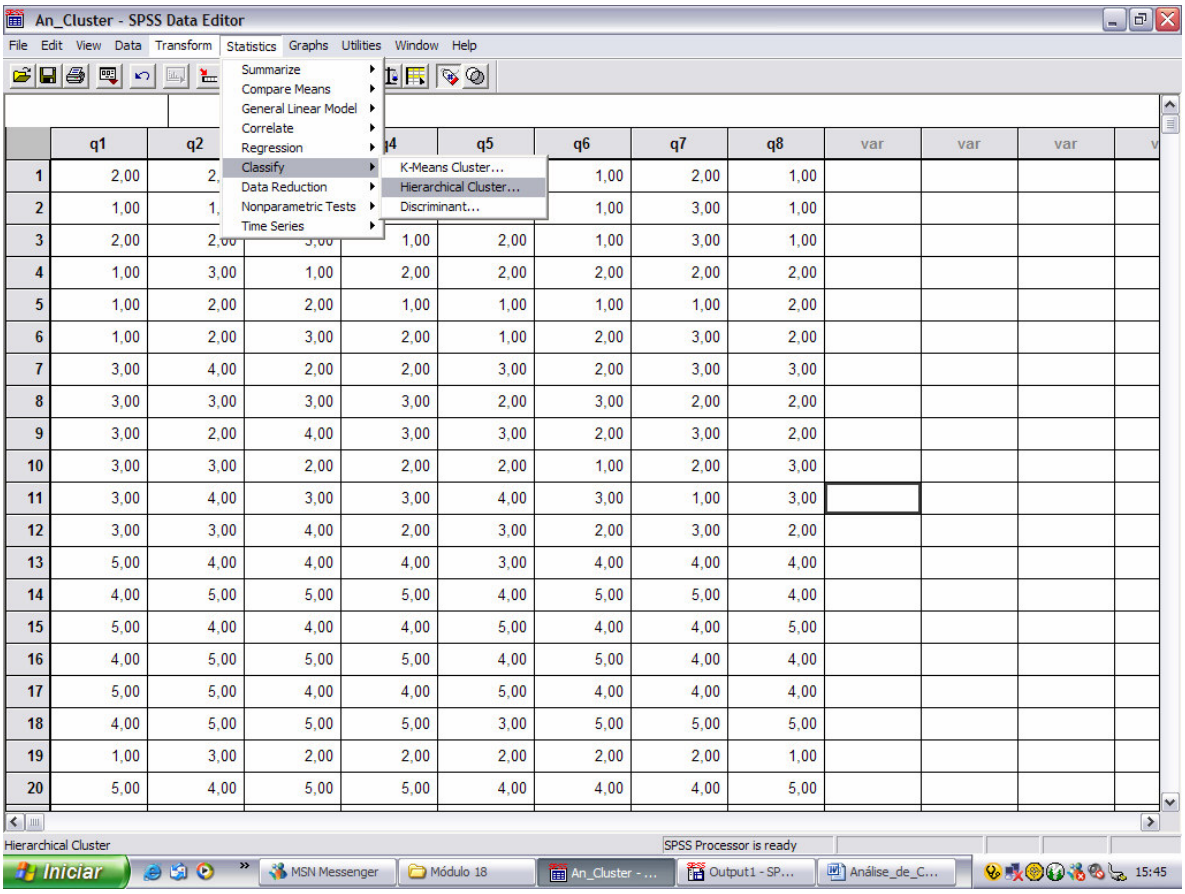
Figura 3: Escolha da ferramenta.

Dentro da opção “classify” estão disponíveis as ferramentas “análise de cluster” e “análise discriminante- Módulo 19”.

Para a análise de cluster, neste momento, devemos escolher o método do agrupamento. Como aprendemos no texto explicativo do módulo, a análise pode seguir o método hierárquico (hierarquical cluster) ou o método não-hierárquico (K-means cluster).

Os métodos não-hierárquicos são menos utilizados, pois demandam grande habilidade do pesquisador que deverá escolher as sementes dos clusters, além de saber previamente quantos clusters deseja obter como resultado.

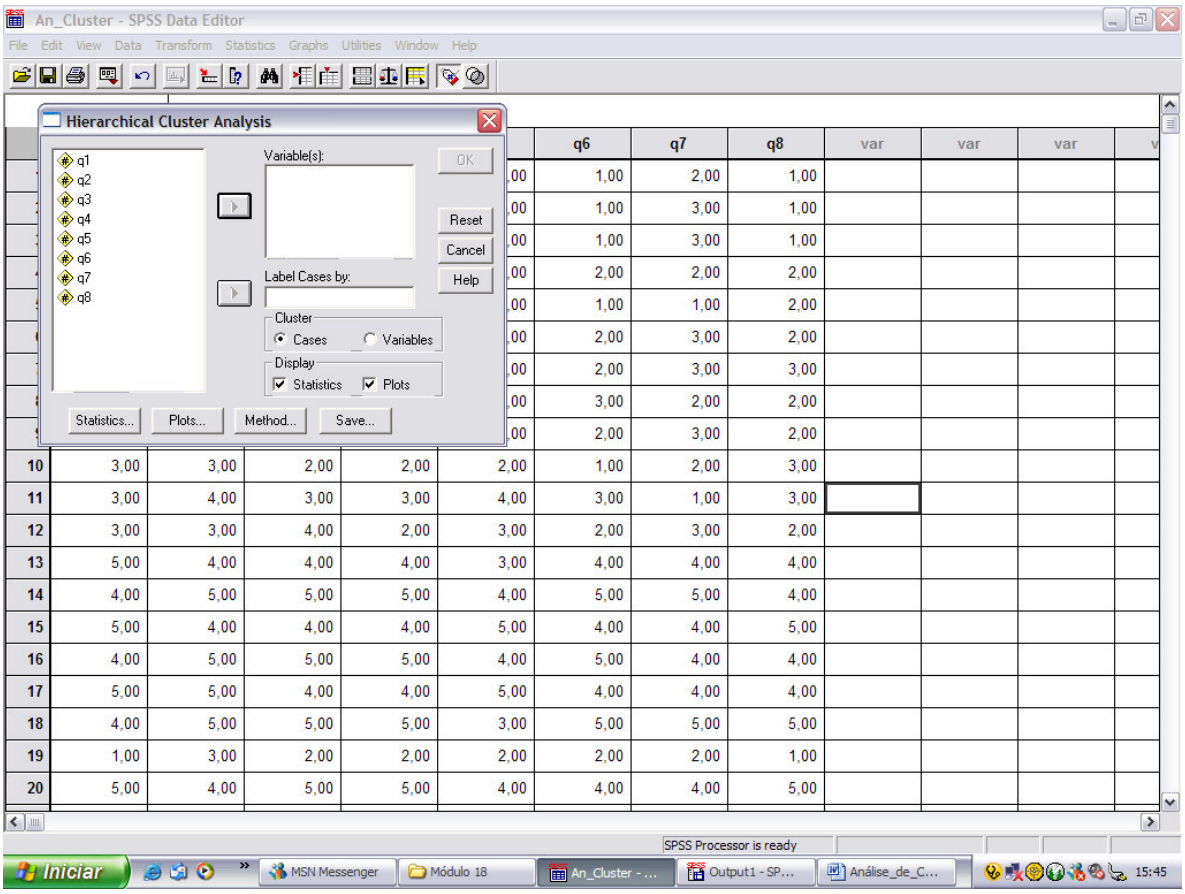
Os métodos hierárquicos são mais populares, portanto devemos selecionar a opção “hierarquical cluster”. No entanto, como no método hierárquico as comparações são feitas par a par, uma base de dados com mais de 500 casos pode levar um tempo de processamento relativamente alto.



	q1	q2			q5	q6	q7	q8	var	var	var
1	2,00	2,				1,00	2,00	1,00			
2	1,00	1,				1,00	3,00	1,00			
3	2,00	2,00		1,00	2,00	1,00	3,00	1,00			
4	1,00	3,00	1,00	2,00	2,00	2,00	2,00	2,00			
5	1,00	2,00	2,00	1,00	1,00	1,00	1,00	2,00			
6	1,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00			
7	3,00	4,00	2,00	2,00	3,00	2,00	3,00	3,00			
8	3,00	3,00	3,00	3,00	2,00	3,00	2,00	2,00			
9	3,00	2,00	4,00	3,00	3,00	2,00	3,00	2,00			
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00			
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00			
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00			
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00			
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00			
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00			
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00			
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00			
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00			
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00			
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00			

Figura 4: Método hierárquico.

Lembre-se que as variáveis devem ser métricas para que a ferramenta possa ser aplicada. É possível aplicar a análise de cluster para variáveis binárias (dicotômicas) e para isto deve-se considerar um outro tipo de medida, que não consiste no enfoque deste tutorial.



						q6	q7	q8	var	var	var
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00			
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00			
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00			
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00			
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00			
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00			
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00			
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00			
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00			
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00			
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00			

Figura 5: Tela inicial de aplicação da ferramenta.

Em “variables” vamos escolher as variáveis segundo as quais os casos (clientes) serão agrupados. Neste problema vamos considerar as respostas dadas pelas clientes às 8 afirmações do questionário, portanto, selecionamos as variáveis “q1; q2; q3; q4; q5; q6; q7; q8”.

Em “label cases by” pode-se selecionar uma variável nominal que nomeia os casos, isto, uma variável que permita saber quem é aquele indivíduo. Por exemplo, poderíamos possuir a variável nome, ou seja, saberíamos o nome da cliente e quais as suas respostas ao questionário. Neste caso a pesquisa foi anônima, ou seja, as clientes não se identificaram no questionário, portanto não selecionaremos nenhuma variável neste menu.

Sabemos que a análise de cluster é utilizada para agrupar casos semelhantes em grupos e os grupos obtidos devem ser distintos entre si. No entanto, esta ferramenta também pode ser utilizada para o agrupamento de variáveis, formando dimensões de variáveis semelhantes

entre si. Existe, porém outra ferramenta mais utilizada para o agrupamento de variáveis, a análise fatorial- módulo 17.

Neste problema vamos agrupar clientes com opiniões semelhantes sobre roupas, então devemos selecionar a opção “cases”.

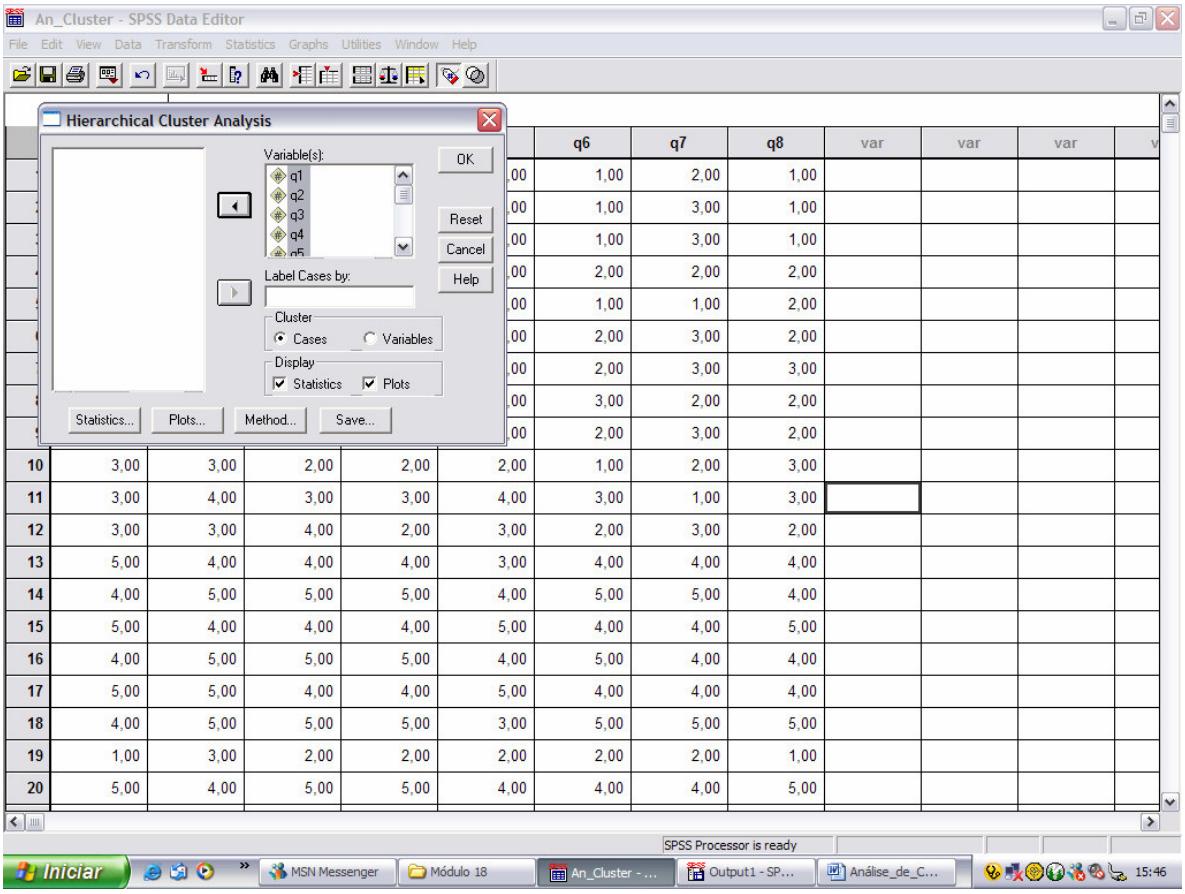


Figura 6: Seleção das variáveis que gerarão o agrupamento.

Em “statistics” devemos selecionar as seguintes opções:

“Agglomeration schedule”: mostra a seqüência de passos na qual o agrupamento foi realizado;

“Proximity matrix”: mostra a matriz de similaridade, isto é, a distância entre os casos;

Em “cluster membership” podemos optar pelo número de clusters desejados. Em “none”, não será apresentado nas tabelas a qual grupo pertence um determinado caso; em “single solution” podemos escolher o número exato de clusters que serão apresentados nas tabelas geradas; em “range of solutions” podemos pedir um intervalo de clusters, por exemplo, se

pedíssemos de 2 a 4 clusters obteríamos os resultados com 2 clusters, 3 clusters e 4 clusters, ou seja, podemos obter a solução com vários números de clusters.

Neste caso, a empresa desejou visualizar três grupos distintos, portanto, escolhemos a opção "single solution" com 3 clusters.

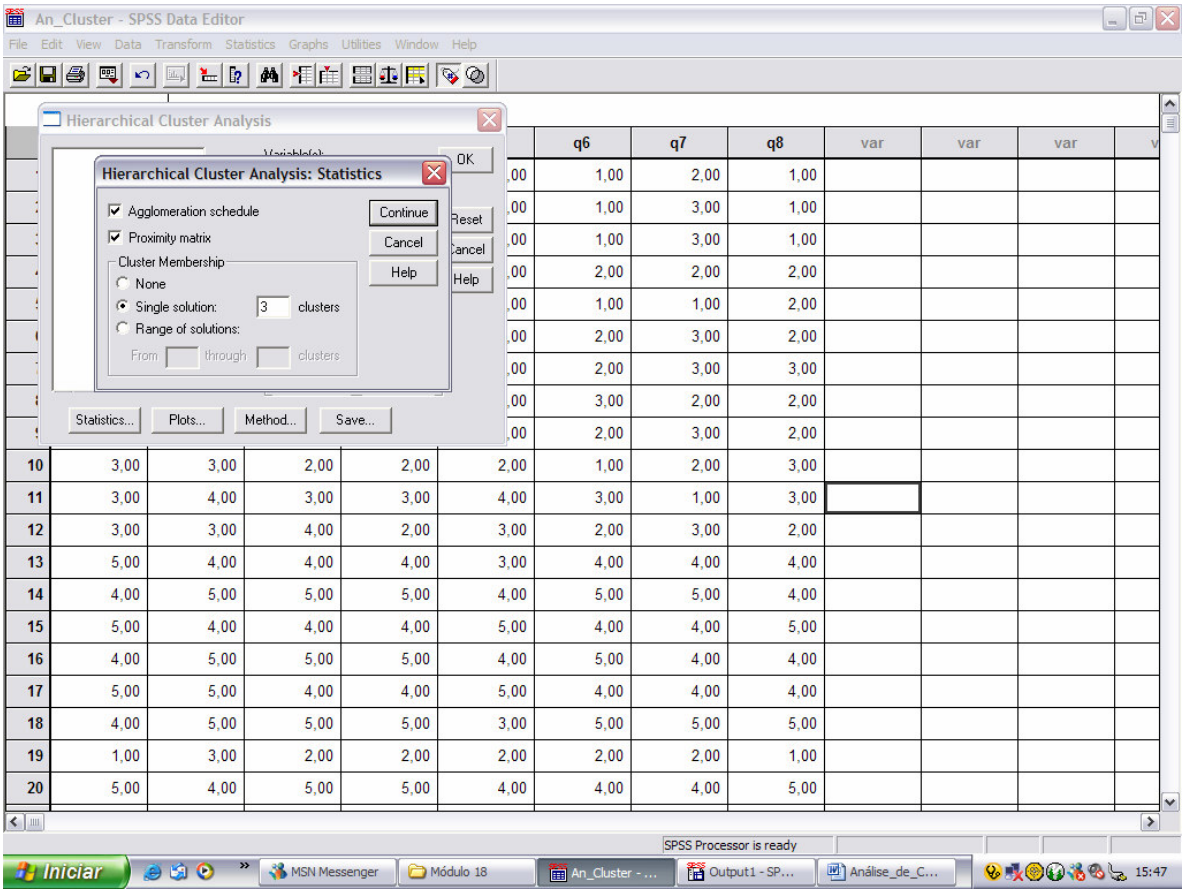


Figura 7: Seleção do número de clusters desejados.

Em "plots" devemos escolher a opção "dendrogram". Esta opção fornece uma representação gráfica da matriz de aglomeração (agglomeration schedule), isto é, mostra como a aglomeração foi realizada, a seqüência como os casos foram agrupados. No quadro "icicle" podemos solicitar uma outra representação gráfica dos clusters construídos, pediremos a representação de todos os clusters encontrados, portanto escolhemos "all clusters".

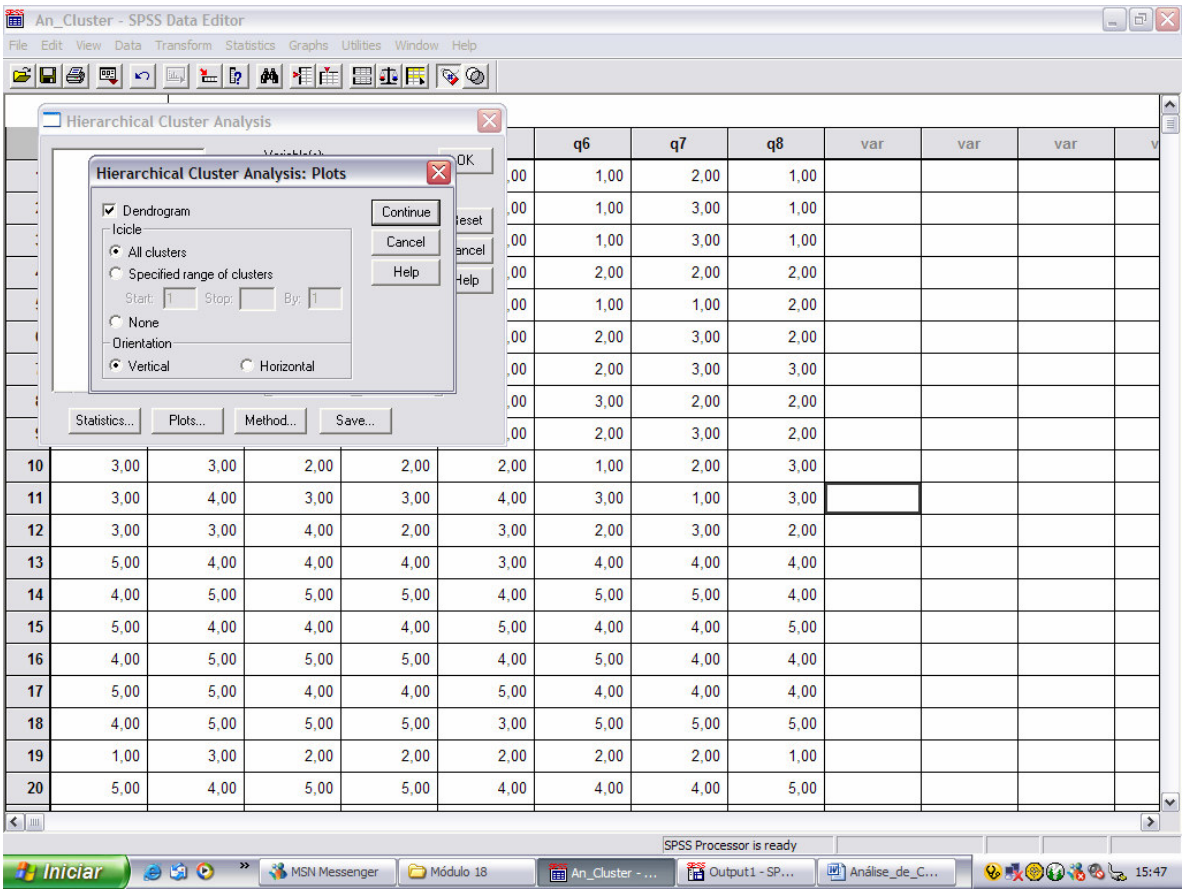


Figura 8: Escolha das representações gráficas.

Em “method” podemos escolher as técnicas de clusterização a serem utilizadas. Em “cluster method” escolheremos qual a medida de distância a ser utilizada para agrupar os casos, todos os métodos apresentados são utilizados, a escolha depende das preferências do pesquisador e das características dos dados estudados. Vamos escolher o método “ward’s method” que mede a distância entre os clusters através da soma dos quadrados entre dois clusters somados para cada variável.

Mais informações sobre os métodos são encontradas no texto explicativo do módulo.

No quadro “measure” podemos escolher a medida de distância a ser utilizada pelo método escolhido (neste caso, pelo método de Ward’s). Lembre-se que medidas de distância diferentes conduzem a resultados diferentes ao agrupamento, assim, aconselha-se utilizar mais de uma medida e comparar os resultados obtidos. Neste caso utilizaremos um tipo de medida mais comumente utilizado: distância euclideana ao quadrado, ou seja, “square euclidean distance”.

No quadro “transformation values” podemos pedir a padronização dos dados. Isto é muito útil quando as variáveis selecionadas para agrupar os casos são medidas em unidades diferentes, por exemplo, se tivéssemos as variáveis: renda- medida em reais, valor da compra- medida em reais, número de compras por ano- medida em unidades; opinião em relação à roupas- medida em notas de 1 a 5 cinco pontos, não poderíamos calcular as distâncias sem padronizar os valores das variáveis, ou haveria séria distorção nos resultados.

A forma mais conhecida de padronização é transformação dos valores em valores padrão através da subtração da média e divisão pelo desvio padrão, assim, cada vetor terá média zero e desvio padrão igual a 1. Para obtermos esta padronização selecionamos “Z scores”.

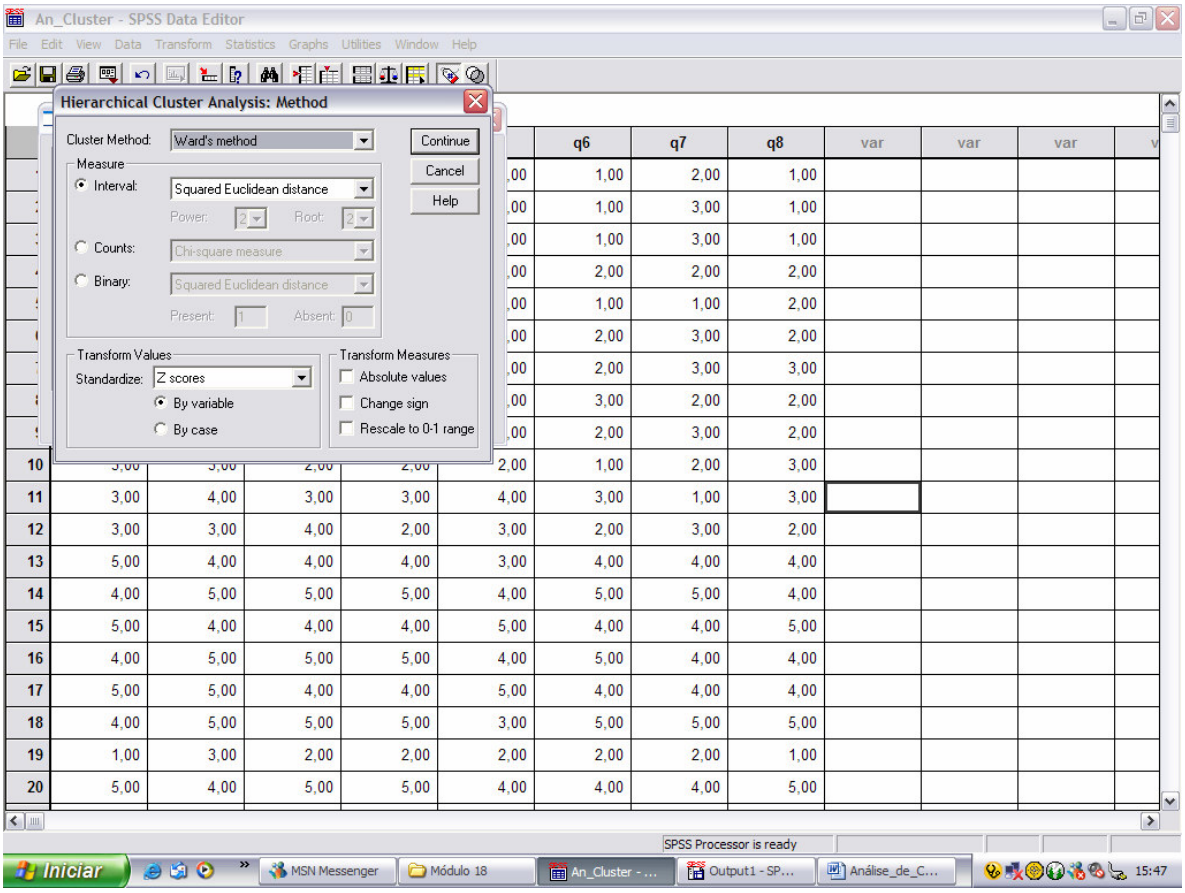


Figura 9: Escolha das técnicas aglomerativas.

Em “save” podemos pedir ao software que crie uma variável que indicará em qual cluster se situará cada um dos casos. Neste problema desejamos obter 3 grupos, então pediremos a

criação da variável com uma solução de três clusters, selecionamos, então, “single solution” com 3 clusters.

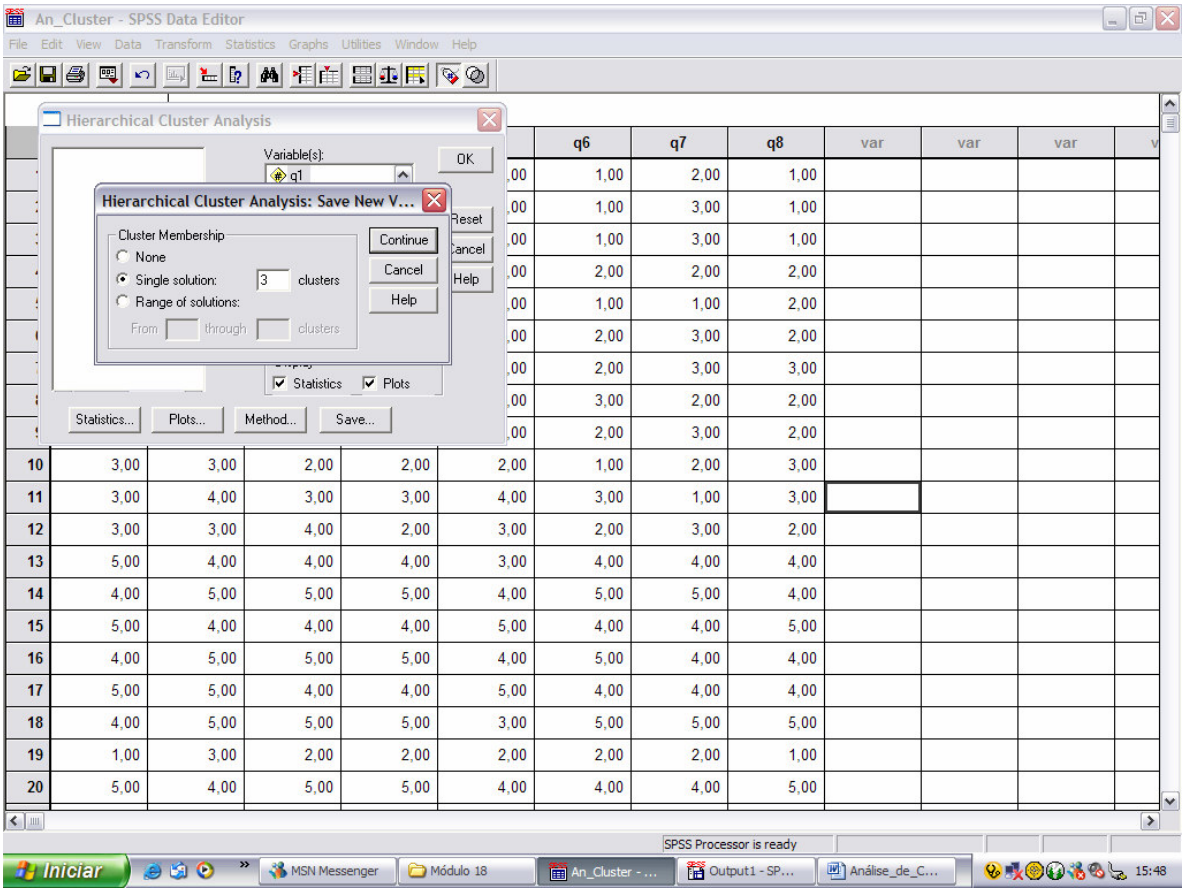
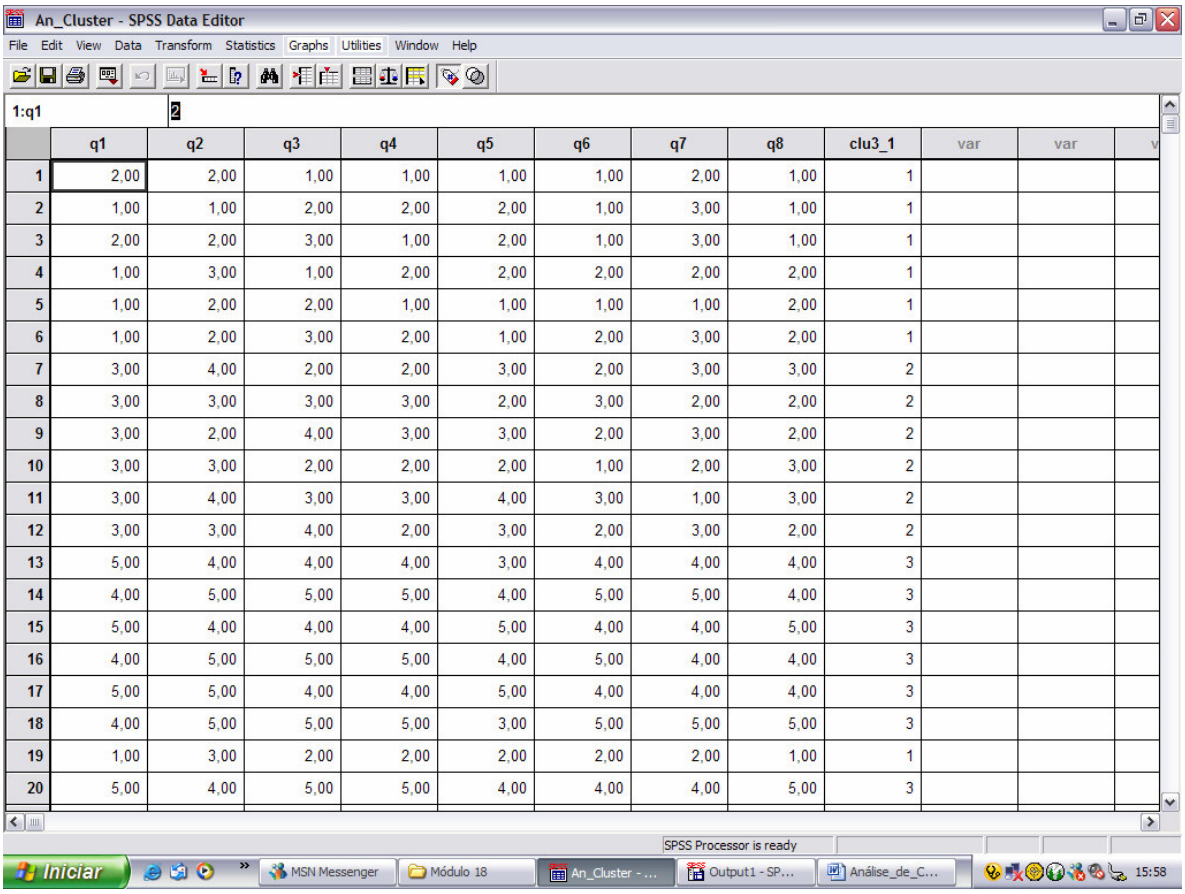


Figura 10: Escolha por se salvar a variável gerada pelo software.

Clique em “ok” para obter as saídas do software.

A planilha inicial ganhou uma nova variável “clu3_1”, que mostra o número do cluster que o caso foi agrupado. Observe a figura 11. Percebemos que os casos 1, 2, 3, 4, 5, 6 e 19 pertencem ao cluster 1; os casos 7, 8, 9, 10, 11, 12 pertencem ao cluster 2 e os casos 13, 14, 15, 16, 17, 18 e 20 pertencem ao cluster 3.



	q1	q2	q3	q4	q5	q6	q7	q8	clu3_1	var	var
1	2,00	2,00	1,00	1,00	1,00	1,00	2,00	1,00	1		
2	1,00	1,00	2,00	2,00	2,00	1,00	3,00	1,00	1		
3	2,00	2,00	3,00	1,00	2,00	1,00	3,00	1,00	1		
4	1,00	3,00	1,00	2,00	2,00	2,00	2,00	2,00	1		
5	1,00	2,00	2,00	1,00	1,00	1,00	1,00	2,00	1		
6	1,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00	1		
7	3,00	4,00	2,00	2,00	3,00	2,00	3,00	3,00	2		
8	3,00	3,00	3,00	3,00	2,00	3,00	2,00	2,00	2		
9	3,00	2,00	4,00	3,00	3,00	2,00	3,00	2,00	2		
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00	2		
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00	2		
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00	2		
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	3		
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00	3		
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00	3		
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00	3		
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00	3		
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00	3		
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00	1		
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00	3		

Figura 11: Planilha com a variável gerada (clu3_1).

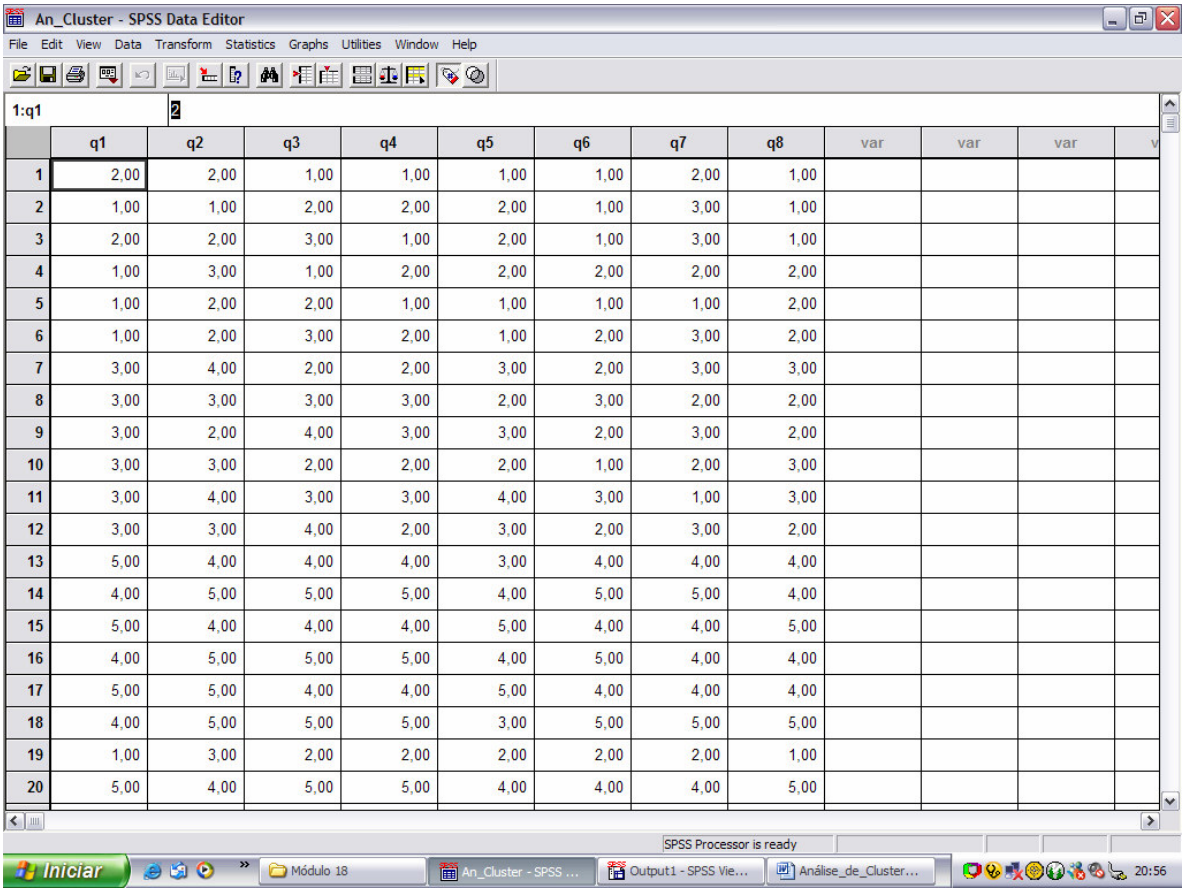
Método não hierárquico

Situação Problema

Utilizaremos agora a mesma situação, mas faremos o agrupamento pelo método não hierárquico.

Preparação dos dados

Os dados devem ser inseridos na planilha como mostra a figura 1.



	q1	q2	q3	q4	q5	q6	q7	q8	var	var	var
1	2,00	2,00	1,00	1,00	1,00	1,00	2,00	1,00			
2	1,00	1,00	2,00	2,00	2,00	1,00	3,00	1,00			
3	2,00	2,00	3,00	1,00	2,00	1,00	3,00	1,00			
4	1,00	3,00	1,00	2,00	2,00	2,00	2,00	2,00			
5	1,00	2,00	2,00	1,00	1,00	1,00	1,00	2,00			
6	1,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00			
7	3,00	4,00	2,00	2,00	3,00	2,00	3,00	3,00			
8	3,00	3,00	3,00	3,00	2,00	3,00	2,00	2,00			
9	3,00	2,00	4,00	3,00	3,00	2,00	3,00	2,00			
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00			
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00			
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00			
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00			
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00			
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00			
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00			
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00			
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00			
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00			
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00			

Figura 1: Inserção dos dados.

Parte 1- Geração de tabelas

Realizaremos uma análise de cluster, ferramenta de agrupamento de casos. Em "statistics" na barra de ferramentas escolhemos a opção "classify".

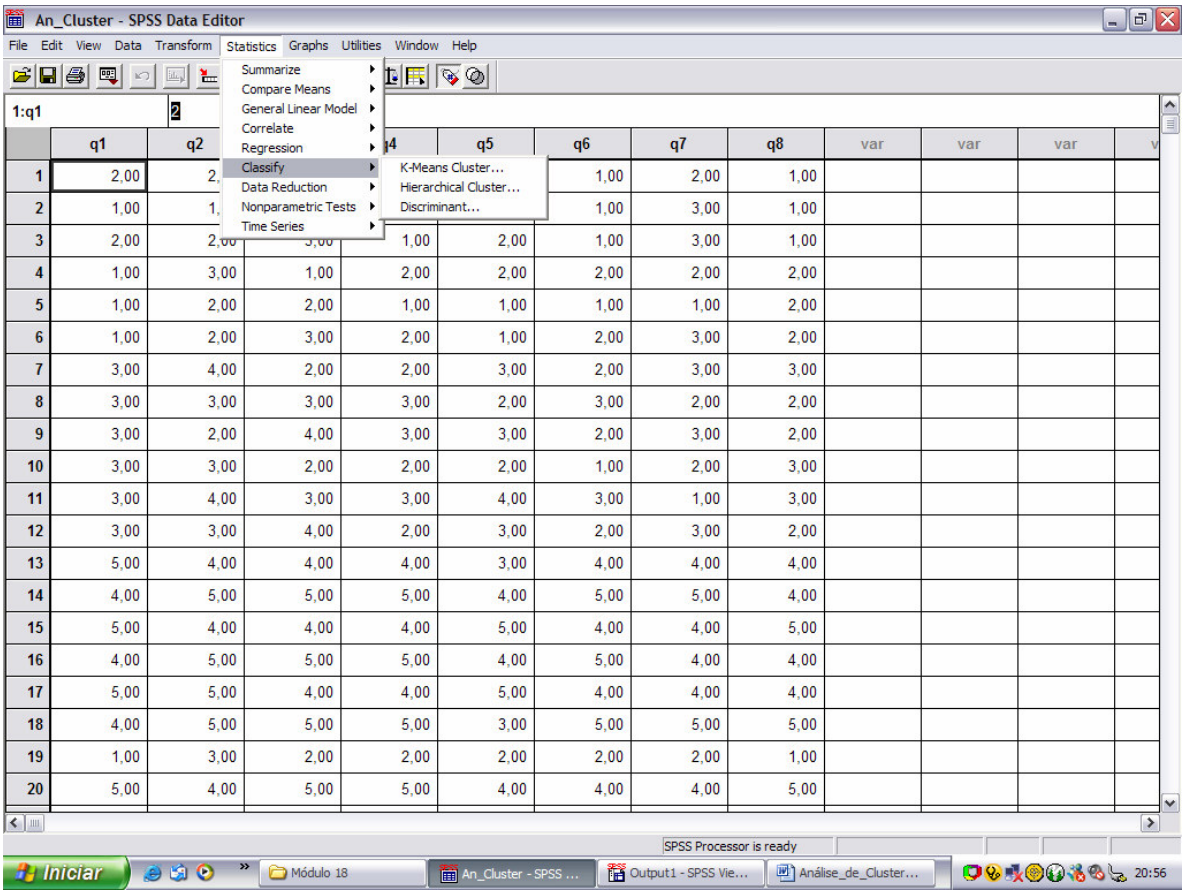


Figura 2: Escolha da ferramenta.

Desejamos agora um agrupamento pelo método não hierárquico. Devemos escolher a opção "K-means cluster".

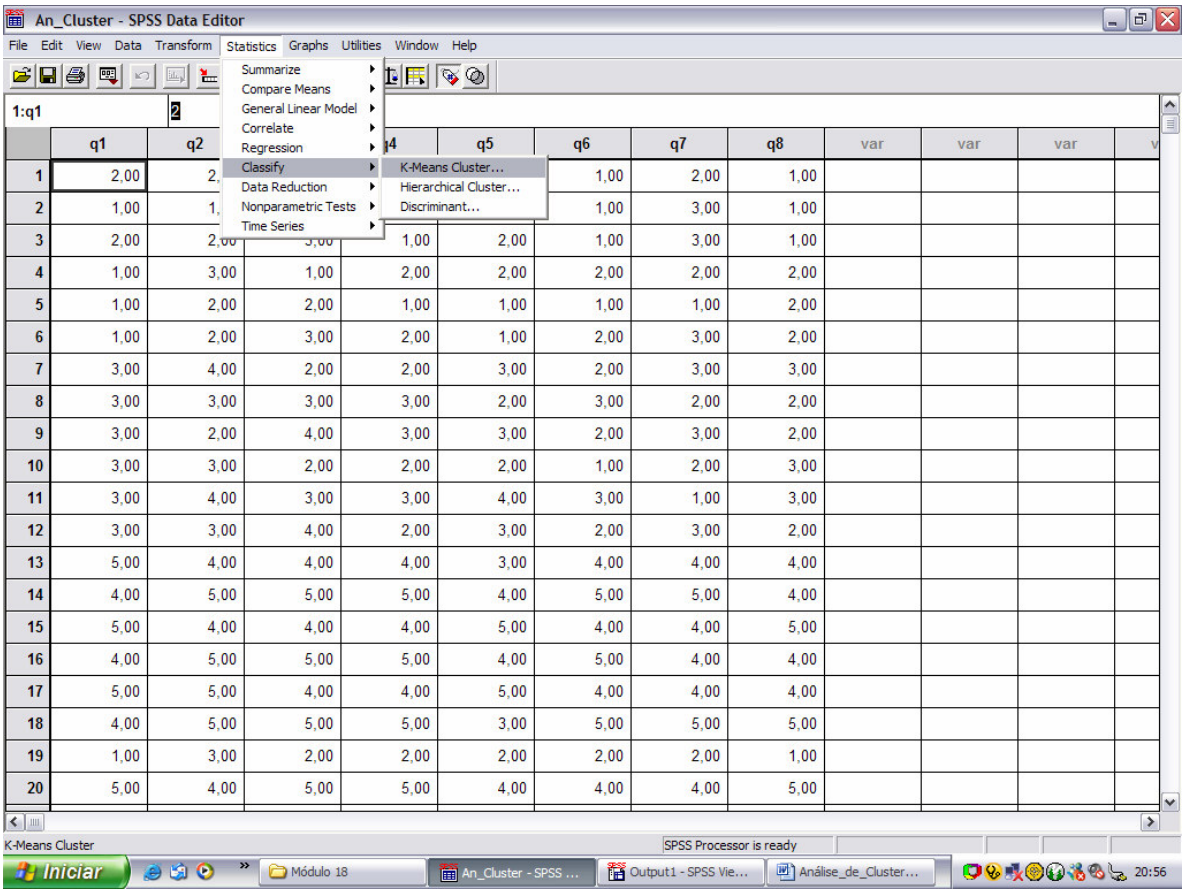


Figura 3: Escolha do método de aglomeração.

Os casos serão agrupados de acordo com as respostas dadas às 8 afirmações do questionário. Lembre-se que as variáveis precisam ser métricas!

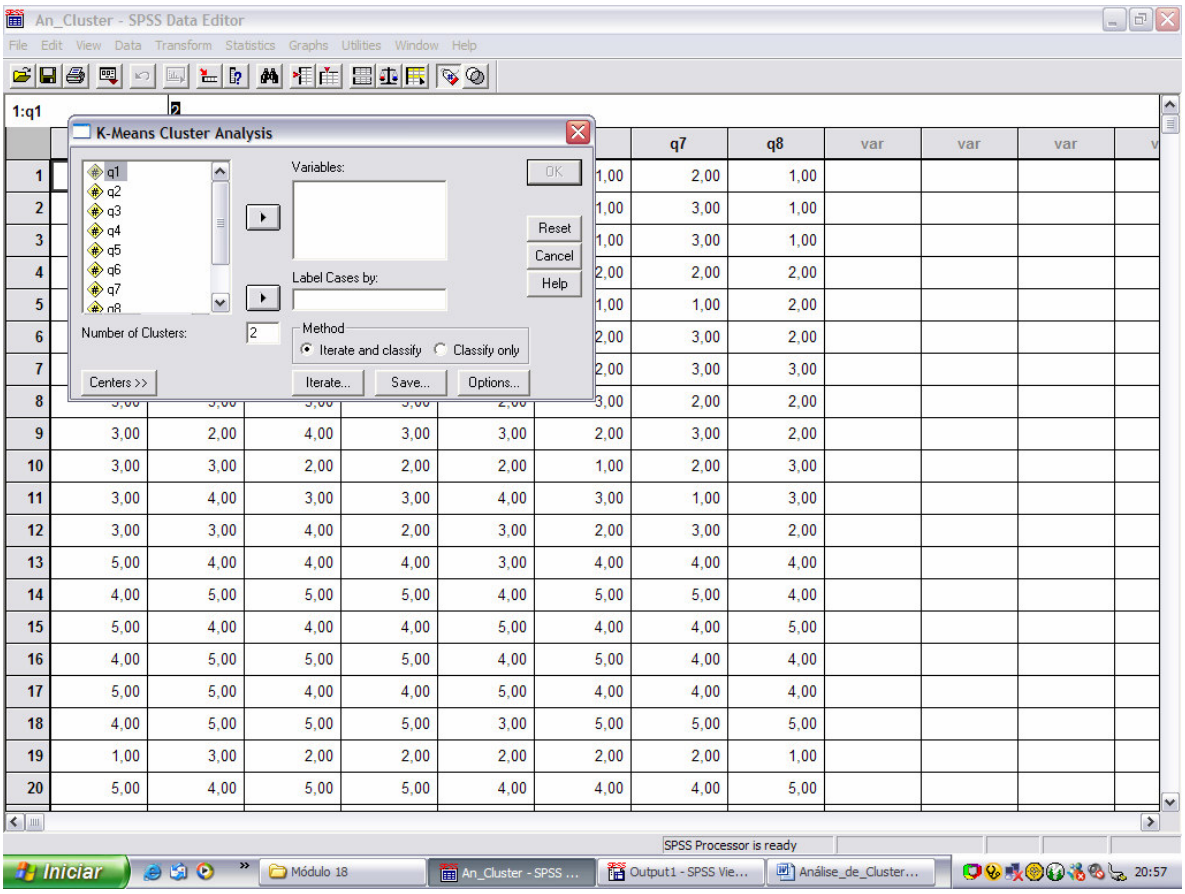


Figura 4: Tela inicial da ferramenta.

Em “variables” escolhemos quais variáveis serão utilizadas para agrupar os casos. Desejamos utilizar as respostas dadas ao questionário, portanto, selecionamos as variáveis “q1, q2, q3, q4, q5, q6, q7, q8”.

Em “label cases by” podemos selecionar uma variável nominal que permita identificar os casos na matriz de similaridades. Neste caso, como a pesquisa beneficiou o anonimato não existe uma variável que permita identificar as clientes que participaram do estudo.

Em “number of clusters” temos que decidir o número de clusters desejados. Note que no método não hierárquico não é possível pedir “intervalos de clusters”, deve-se decidir quantos clusters exatamente devem ser formados.

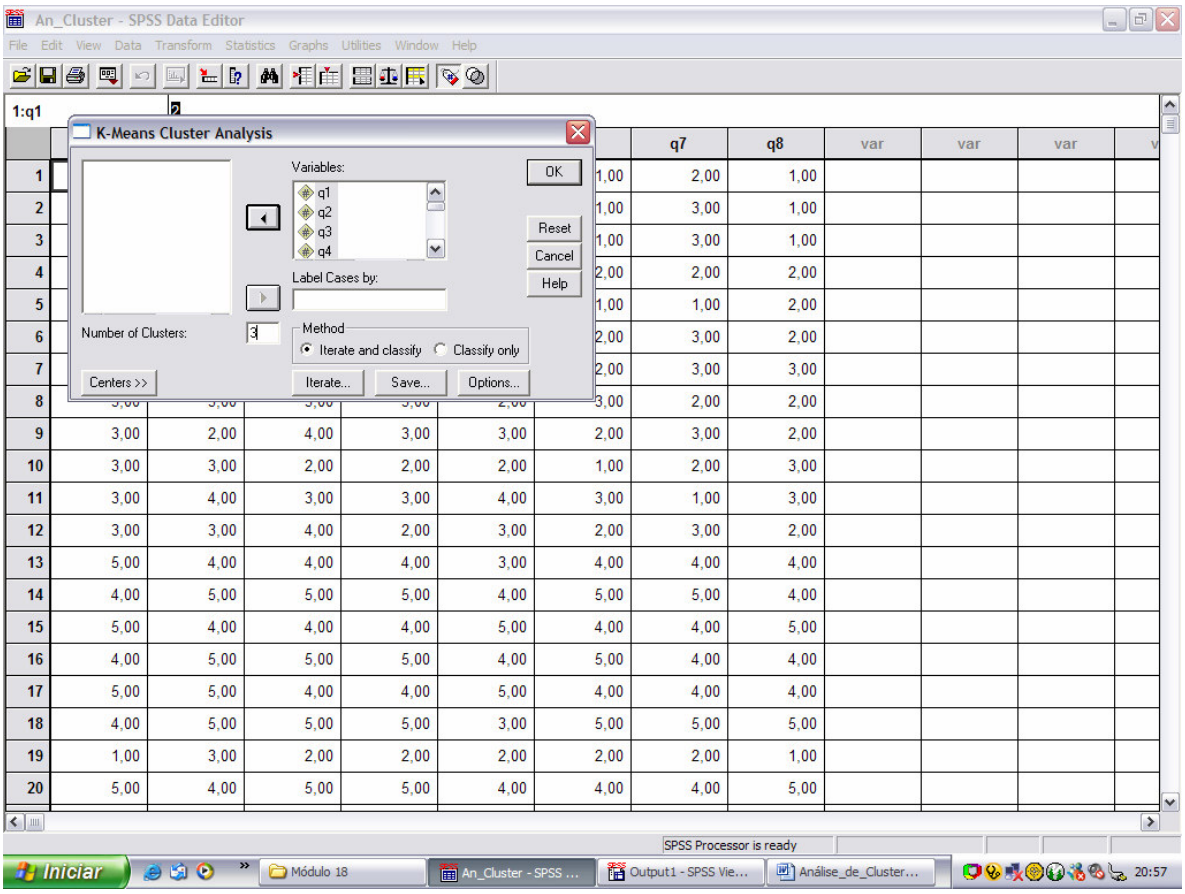


Figura 5: Seleção das variáveis e do número de clusters.

No menu “iterate” temos:

“Maximum interations”: limita o número de interações do algoritmo K-means. Caso o critério de convergência não seja satisfeito e o número de interações alcançar o valor máximo estabelecido esta será a solução final apresentada. O número de interações pode ser qualquer valor entre 1 e 999. Neste caso escolheremos 30 interações;

“Convergence criterion”: determina quando as interações devem parar. Este valor representa uma proporção da mínima distância entre os centros iniciais dos clusters. Deve ser um valor maior que zero e menor que 1. Por exemplo, se o critério for igual a 0,02, a interação cessaria quando uma interação completa não move nenhum centro de cluster por uma distância maior que 0,02 (2%) da menor distância entre qualquer distância dos centros dos clusters iniciais;

“Use running means”: permite que os centros dos clusters sejam recalculados depois que cada caso é agrupado. Não selecionaremos esta opção.

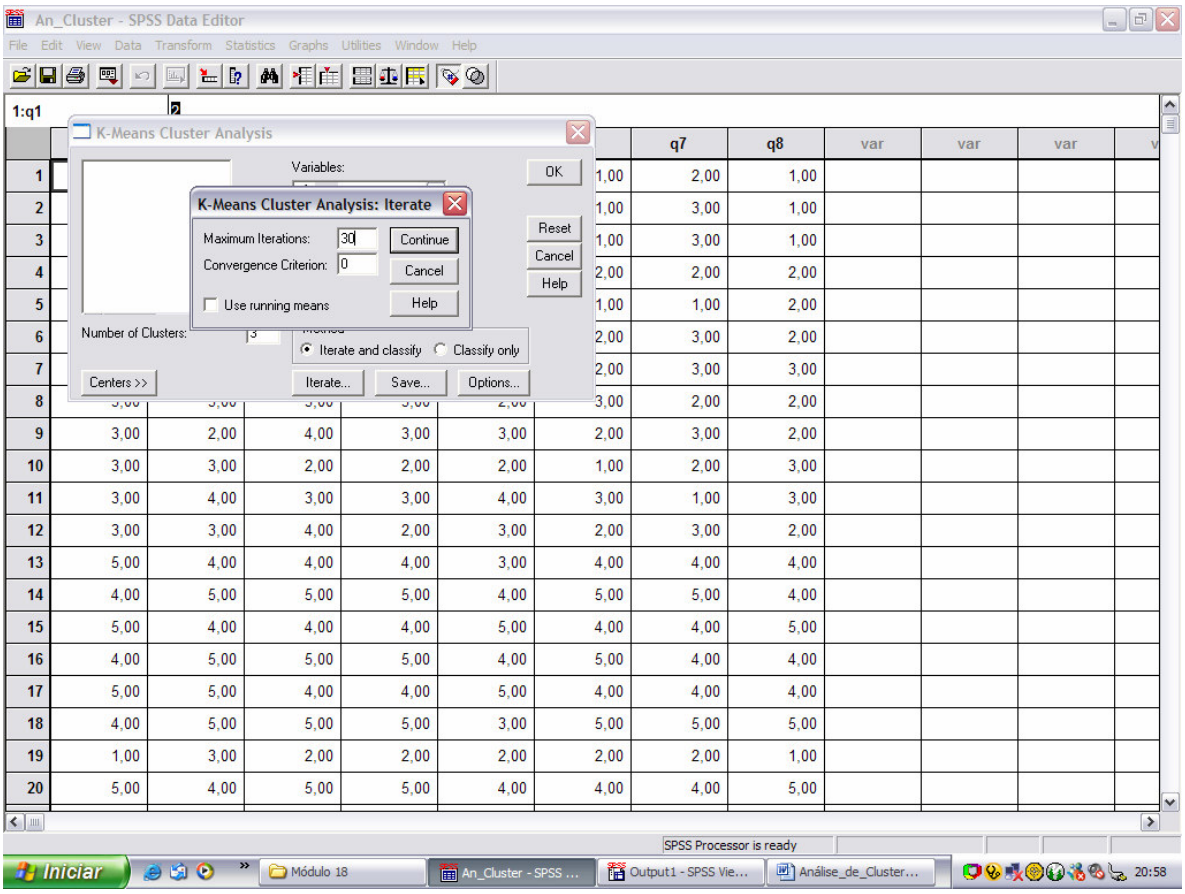


Figura 6: Escolha do número de interações.

No menu “save” podemos pedir ao software que gere variáveis. Selecionaremos a opção “cluster membership” que dará a variável “qcl_1”- mostra a qual cluster pertence cada caso. Selecionaremos a opção “distance from cluster center” que dará a variável “qcl_2”- mostra a distância de cada caso ao centro do seu cluster.

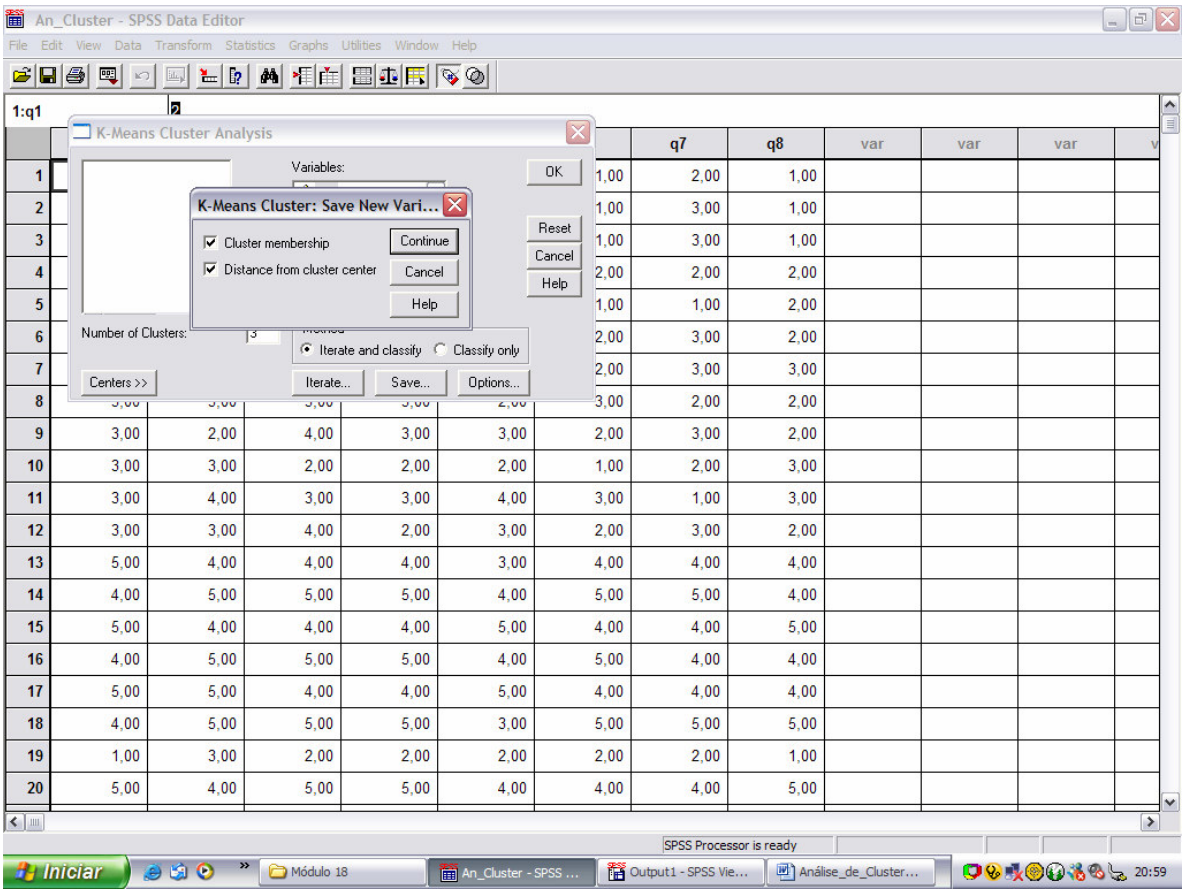


Figura 7: Escolha das variáveis geradas.

No menu “options” escolhemos saídas adicionais do software. Em “statistics” escolhemos a opção “initial clusters centers”, que dará as sementes iniciais escolhidas pelo próprio software para iniciar a análise (lembre-se que um requisito do método não hierárquico é a escolha de sementes iniciais!); a opção “ANOVA table”, que mostra a análise de variância para cada variável; a opção “cluster information for each case”, que mostra em qual cluster está cada caso e qual a distância de cada um do centro do seu cluster.

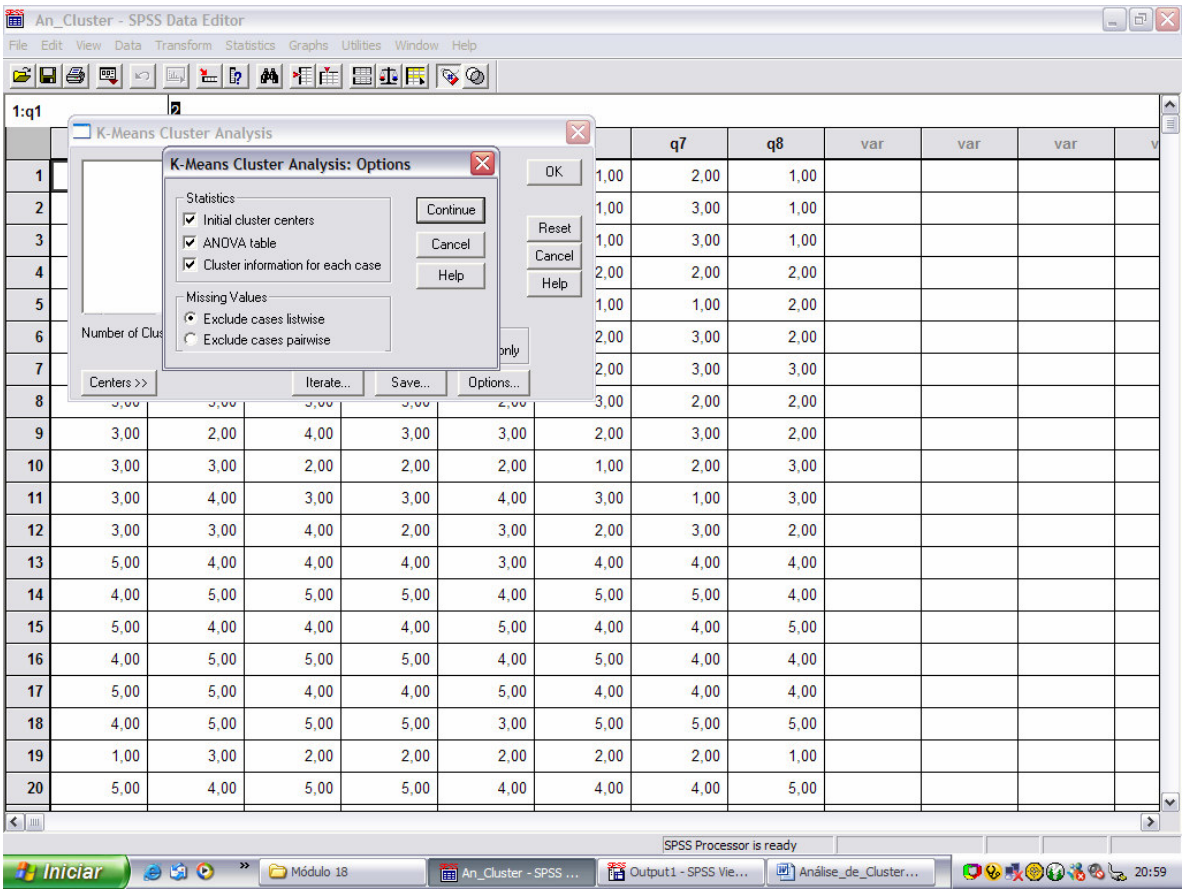
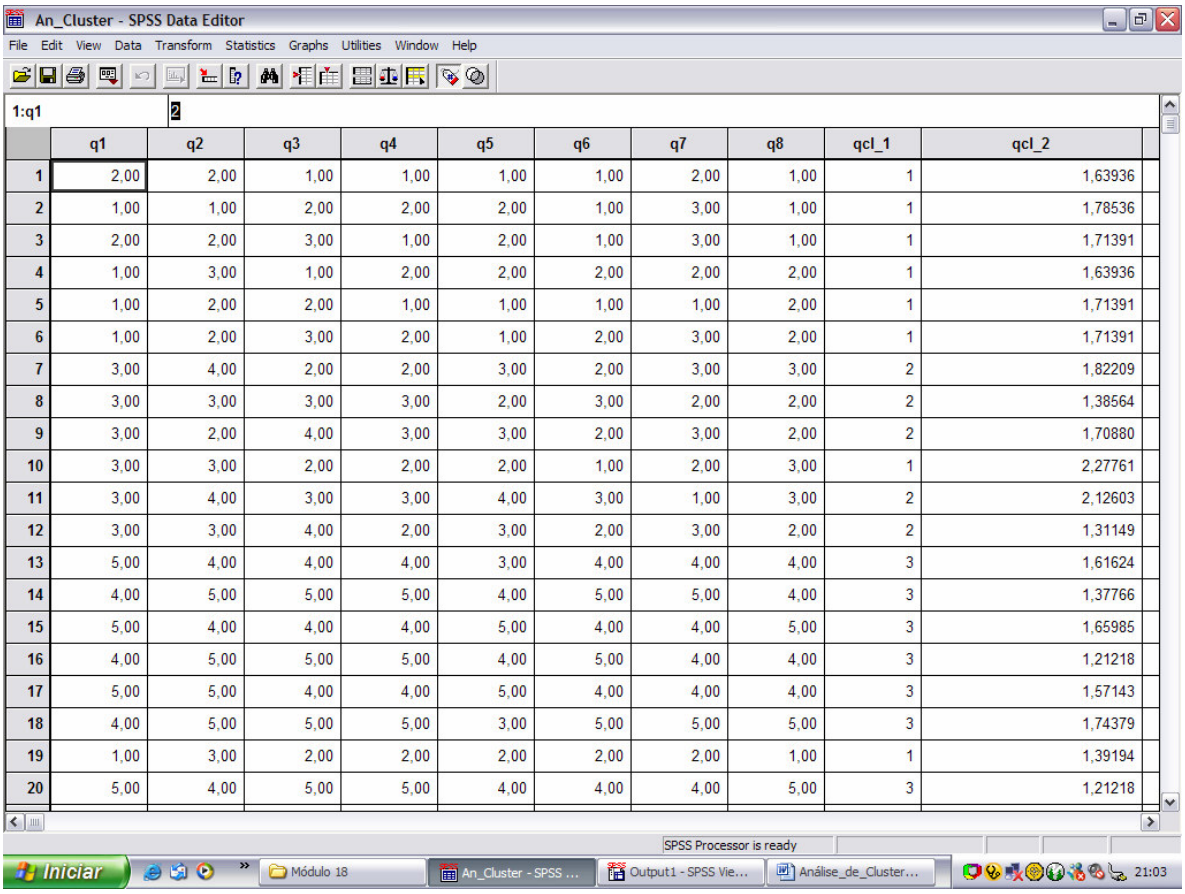


Figura 8: Escolha das estatísticas desejadas.

Clique em “ok” para obter os resultados da análise.

Observe que após a análise temos duas novas variáveis na planilha de dados. A variável qcl_1 mostra que os casos 1, 2, 3, 4, 5, 6, 10 e 19 pertencem ao cluster 1; os casos 7, 8, 9, 11, 12 pertencem ao cluster 2; os casos 13, 14, 15, 16, 17, 18, 20 pertencem ao cluster 3.
A variável qcl_2 mostra distância de cada caso ao centro do seu cluster. Por exemplo, o caso 1, que pertence ao cluster 1 está a uma distância de 1,63936 do centro do cluster 1.



	q1	q2	q3	q4	q5	q6	q7	q8	qcl_1	qcl_2
1	2,00	2,00	1,00	1,00	1,00	1,00	2,00	1,00	1	1,63936
2	1,00	1,00	2,00	2,00	2,00	1,00	3,00	1,00	1	1,78536
3	2,00	2,00	3,00	1,00	2,00	1,00	3,00	1,00	1	1,71391
4	1,00	3,00	1,00	2,00	2,00	2,00	2,00	2,00	1	1,63936
5	1,00	2,00	2,00	1,00	1,00	1,00	1,00	2,00	1	1,71391
6	1,00	2,00	3,00	2,00	1,00	2,00	3,00	2,00	1	1,71391
7	3,00	4,00	2,00	2,00	3,00	2,00	3,00	3,00	2	1,82209
8	3,00	3,00	3,00	3,00	2,00	3,00	2,00	2,00	2	1,38564
9	3,00	2,00	4,00	3,00	3,00	2,00	3,00	2,00	2	1,70880
10	3,00	3,00	2,00	2,00	2,00	1,00	2,00	3,00	1	2,27761
11	3,00	4,00	3,00	3,00	4,00	3,00	1,00	3,00	2	2,12603
12	3,00	3,00	4,00	2,00	3,00	2,00	3,00	2,00	2	1,31149
13	5,00	4,00	4,00	4,00	3,00	4,00	4,00	4,00	3	1,61624
14	4,00	5,00	5,00	5,00	4,00	5,00	5,00	4,00	3	1,37766
15	5,00	4,00	4,00	4,00	5,00	4,00	4,00	5,00	3	1,65985
16	4,00	5,00	5,00	5,00	4,00	5,00	4,00	4,00	3	1,21218
17	5,00	5,00	4,00	4,00	5,00	4,00	4,00	4,00	3	1,57143
18	4,00	5,00	5,00	5,00	3,00	5,00	5,00	5,00	3	1,74379
19	1,00	3,00	2,00	2,00	2,00	2,00	2,00	1,00	1	1,39194
20	5,00	4,00	5,00	5,00	4,00	4,00	4,00	5,00	3	1,21218

Figura 9: Variáveis criadas pelo software.