

Módulo 19 - Análise Discriminante

Geração de Tabelas

Situação Problema

Um banco deseja classificar seus clientes de acordo com seu perfil de investimento: investidor conservador, investidor moderado ou investidor arrojado. Sabe-se que este perfil depende da renda do investidor (mensal em reais), do valor de seus investimentos (total acumulado em reais) e do tempo que ele é cliente do banco (em meses). O investidor conservador é aquele avesso aos riscos, que investe em modalidades de baixo risco, mesmo que estas ofereçam baixa remuneração ao capital; os moderados são aqueles que não são avessos ao risco, mas também não privilegiam apenas investimentos de risco elevado, fazem alguns investimentos de risco elevado e outros de baixo risco, ou preferem investir em modalidades de risco médio; os arrojados são os investidores que possuem pouca aversão ao risco, preferem investimentos arriscados com alta remuneração do capital.

O banco, então, montou um banco de dados com estas informações de 60 clientes, e deseja, a partir deles obter uma equação que permita classificar casos futuros nas categorias propostas (investidor conservador, moderador e arrojado).

A técnica indicada para resolver este problema é a análise discriminante.

Preparação dos dados

A análise discriminante é uma técnica de dependência. Possui uma variável dependente categórica e uma ou mais variáveis independentes métricas.

Esta técnica tenta classificar os casos, a partir da relação entre as variáveis independentes e a dependente, dentro das categorias específicas, que neste caso são investidor conservador, moderado ou arrojado.

Aprendemos no texto explicativo do módulo que é importante para utilização da análise discriminante dividir a amostra em duas partes, a amostra de construção com a qual se construirá o modelo e a amostra de validação, com a qual se valida o modelo.

Devemos, então, inserir uma variável, ademais das variáveis dependente e independentes, possibilitando selecionar os casos para construção e validação. Chamamos esta variável de “seleção”. Esta variável terá duas categorias: uma irá

Módulo 19 – Análise Discriminante

representar os casos que serão utilizados para construir a função (denominada “amostra”) e a outra servirá para validação da função construída (denominada “validação”).

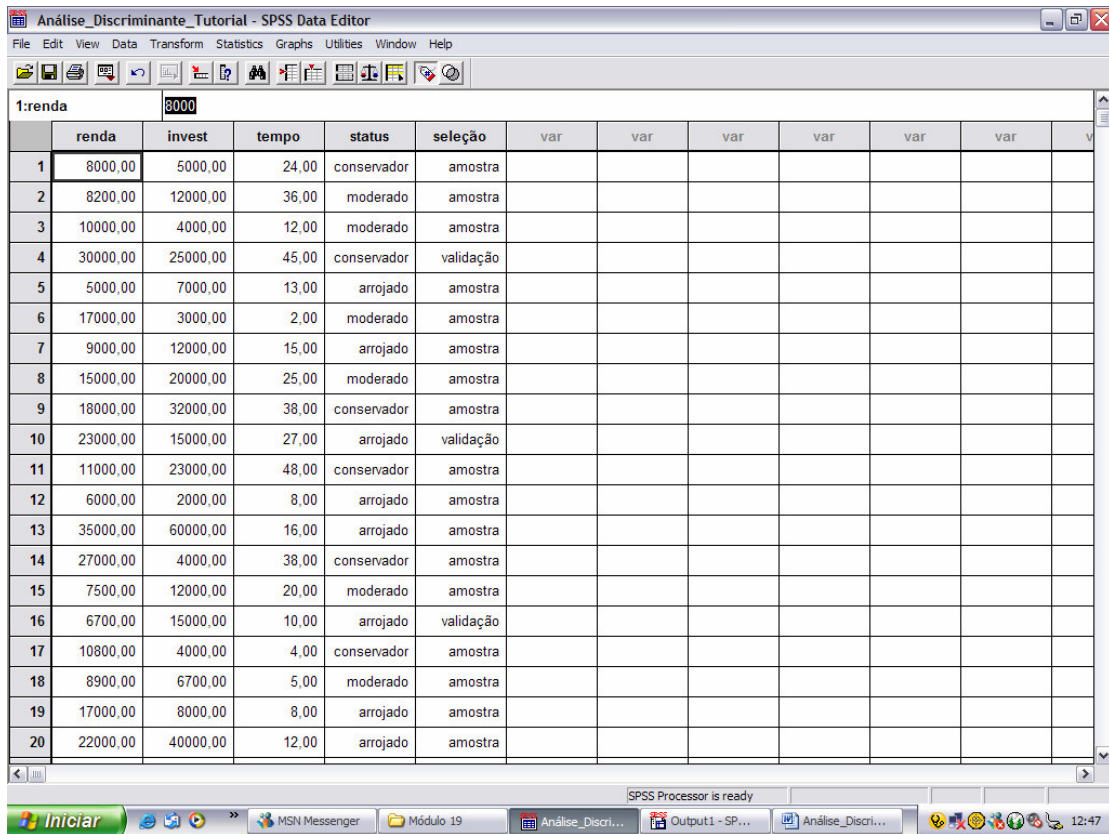
Uma observação importante é quanto ao número de casos necessários. De acordo com a literatura, deve-se utilizar no mínimo 20 casos para cada categoria da variável dependente. Neste caso como temos 3 categorias (conservadores, moderados, arrojados) seriam necessários 60 casos.

Dividimos os casos da seguinte forma para validação do modelo:

Casos	Conservadores	Moderados	Arrojados
Construção	14	14	14
Validação	6	6	6

Quadro 1: Casos para validação do modelo.

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



	renda	invest	tempo	status	seleção	var	var	var	var	var	var	var
1	8000,00	5000,00	24,00	conservador	amostra							
2	8200,00	12000,00	36,00	moderado	amostra							
3	10000,00	4000,00	12,00	moderado	amostra							
4	30000,00	25000,00	45,00	conservador	validação							
5	5000,00	7000,00	13,00	arrojado	amostra							
6	17000,00	3000,00	2,00	moderado	amostra							
7	9000,00	12000,00	15,00	arrojado	amostra							
8	15000,00	20000,00	25,00	moderado	amostra							
9	18000,00	32000,00	38,00	conservador	amostra							
10	23000,00	15000,00	27,00	arrojado	validação							
11	11000,00	23000,00	48,00	conservador	amostra							
12	6000,00	2000,00	8,00	arrojado	amostra							
13	35000,00	60000,00	16,00	arrojado	amostra							
14	27000,00	4000,00	38,00	conservador	amostra							
15	7500,00	12000,00	20,00	moderado	amostra							
16	6700,00	15000,00	10,00	arrojado	validação							
17	10800,00	4000,00	4,00	conservador	amostra							
18	8900,00	6700,00	5,00	moderado	amostra							
19	17000,00	8000,00	8,00	arrojado	amostra							
20	22000,00	40000,00	12,00	arrojado	amostra							

Figura 1: Inserção dos dados.

Parte 1- Geração de tabelas

A análise discriminante é uma técnica classificatória, isto é, tentará classificar os casos em categorias, portanto, na barra de ferramentas clicaremos em “statistics” e escolheremos a opção “classify”.

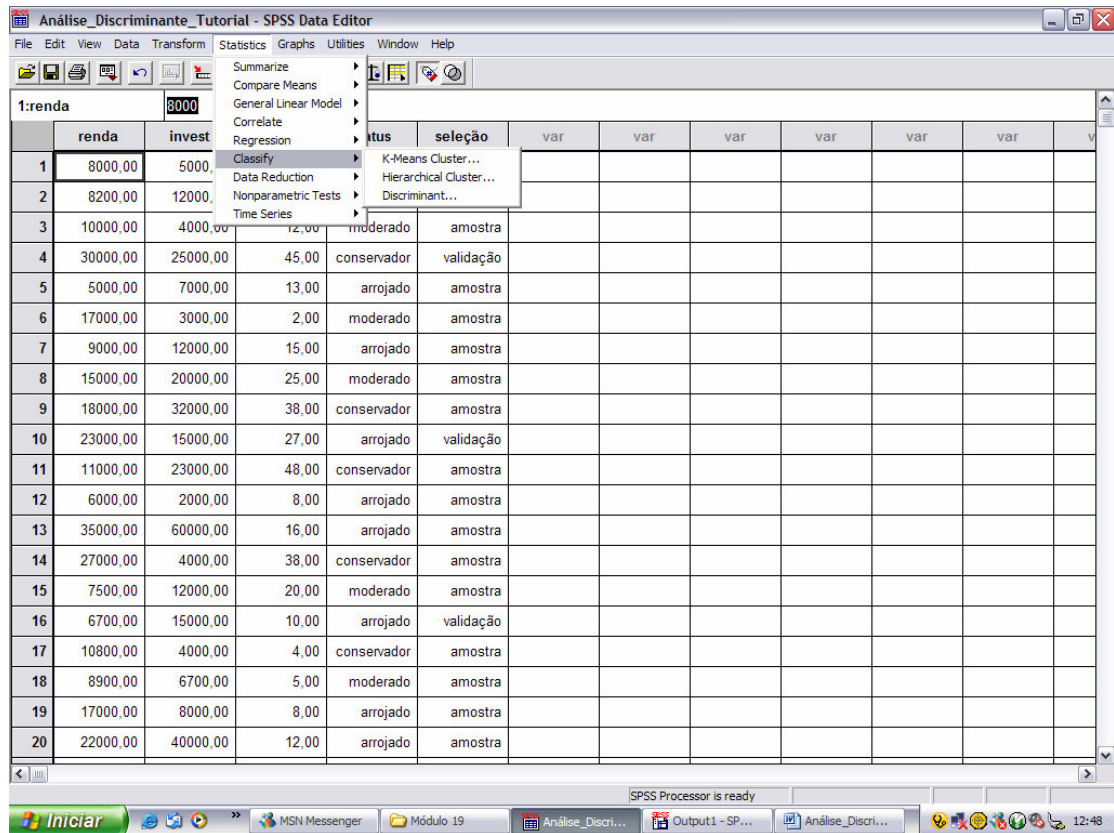
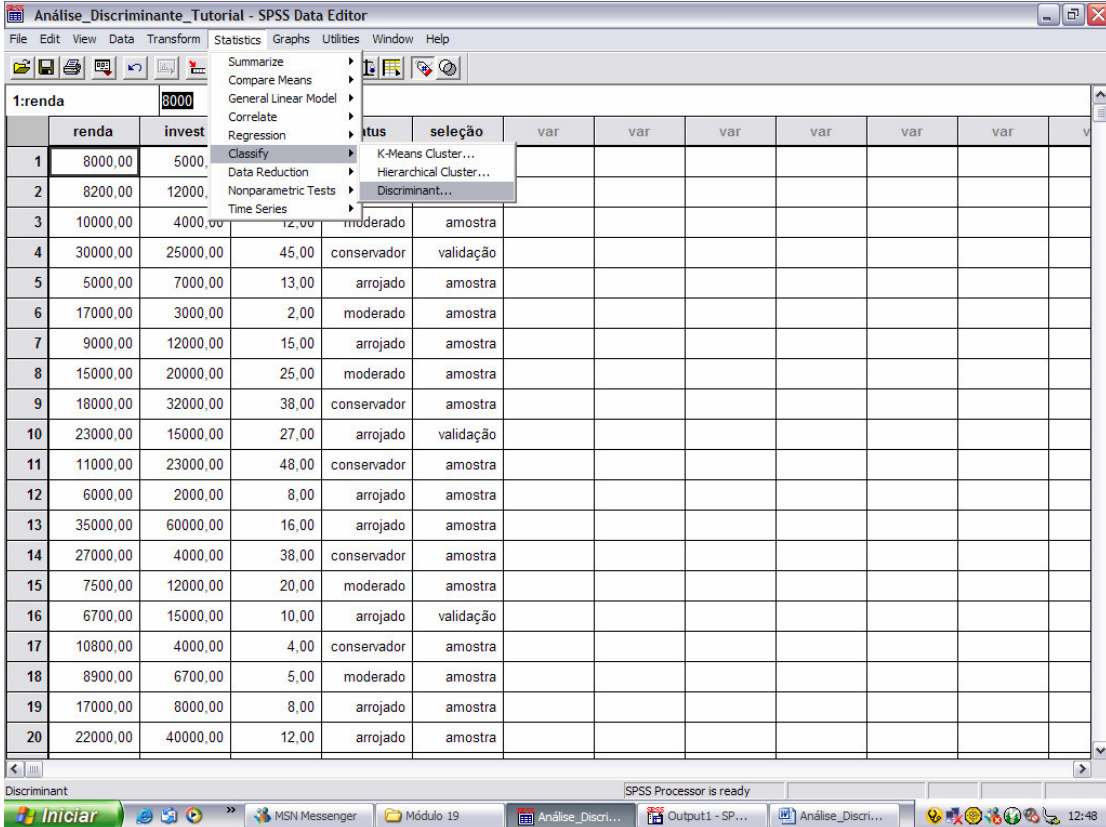


Figura 2: Escolha da ferramenta.

Neste problema queremos construir uma função de classificação para classificar os clientes de um determinado banco em categorias de investidor pré-estabelecidas, ou seja, já conhecemos os possíveis grupos que existirão para construir a função (ou funções) que poderá ser utilizada para avaliar novos casos. Escolhemos então “discriminant”.

Módulo 19 – Análise Discriminante



The screenshot shows the SPSS Data Editor window titled "Análise_Discriminante_Tutorial - SPSS Data Editor". The menu bar includes File, Edit, View, Data, Transform, Statistics, Graphs, Utilities, Window, and Help. The "Statistics" menu is open, and the "Classify" option is selected, which has opened a sub-menu. In this sub-menu, "Discriminant..." is highlighted. The main data grid contains 20 rows of data with columns for "renda", "invest", "atuz", and "seleção". The "seleção" column contains values like "amostra" and "validação". The taskbar at the bottom shows the Windows Start button, "Iniciar", and several open applications including "Módulo 19", "Análise_Discri...", "Output1 - SP...", and "Análise_Discri...". The system clock shows 12:48.

	renda	invest	atuz	seleção	var	var	var	var	var	var	var	var
1	8000,00	5000,00										
2	8200,00	12000,00										
3	10000,00	4000,00	12,00	moderado	amostra							
4	30000,00	25000,00	45,00	conservador	validação							
5	5000,00	7000,00	13,00	arrojado	amostra							
6	17000,00	3000,00	2,00	moderado	amostra							
7	9000,00	12000,00	15,00	arrojado	amostra							
8	15000,00	20000,00	25,00	moderado	amostra							
9	18000,00	32000,00	38,00	conservador	amostra							
10	23000,00	15000,00	27,00	arrojado	validação							
11	11000,00	23000,00	48,00	conservador	amostra							
12	6000,00	2000,00	8,00	arrojado	amostra							
13	35000,00	60000,00	16,00	arrojado	amostra							
14	27000,00	4000,00	38,00	conservador	amostra							
15	7500,00	12000,00	20,00	moderado	amostra							
16	6700,00	15000,00	10,00	arrojado	validação							
17	10800,00	4000,00	4,00	conservador	amostra							
18	8900,00	6700,00	5,00	moderado	amostra							
19	17000,00	8000,00	8,00	arrojado	amostra							
20	22000,00	40000,00	12,00	arrojado	amostra							

Figura 3: Seleção da análise discriminante.

Módulo 19 – Análise Discriminante

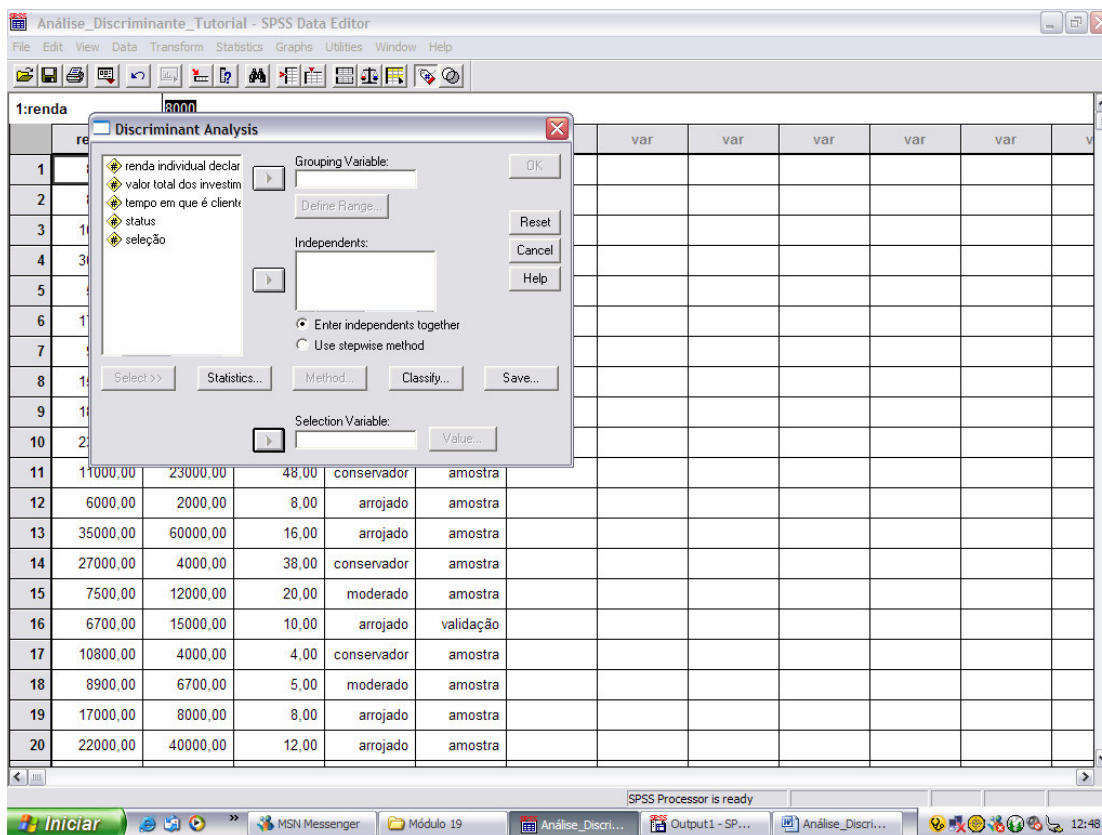


Figura 4: Tela inicial de aplicação da ferramenta.

A variável que representa o status de cada caso nos grupos preestabelecidos e segundo a qual o agrupamento será feito é a variável dependente, neste caso denominada "status". Em "grouping variable" selecionamos "status".

Módulo 19 – Análise Discriminante

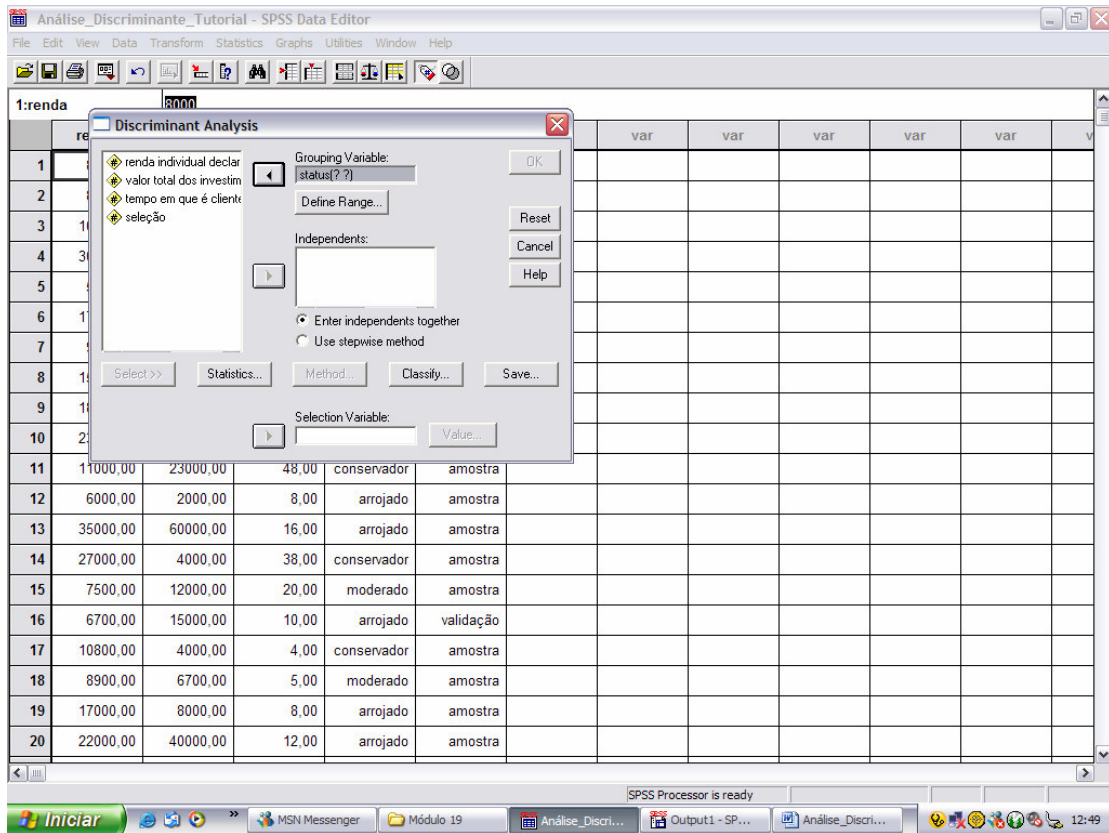


Figura 5: Seleção da variável dependente.

Trata-se de uma variável categórica e nesta ferramenta podemos decidir quantas das categorias serão consideradas na análise. Em “define range” decidiremos de qual a qual categoria incluiremos na análise. Neste problema temos a seguinte classificação:

0= conservador;

1= moderado;

2= arrojado.

Como desejamos classificar os clientes nas três categorias colocaremos “minimum=0” e “maximum=2”.

Módulo 19 – Análise Discriminante

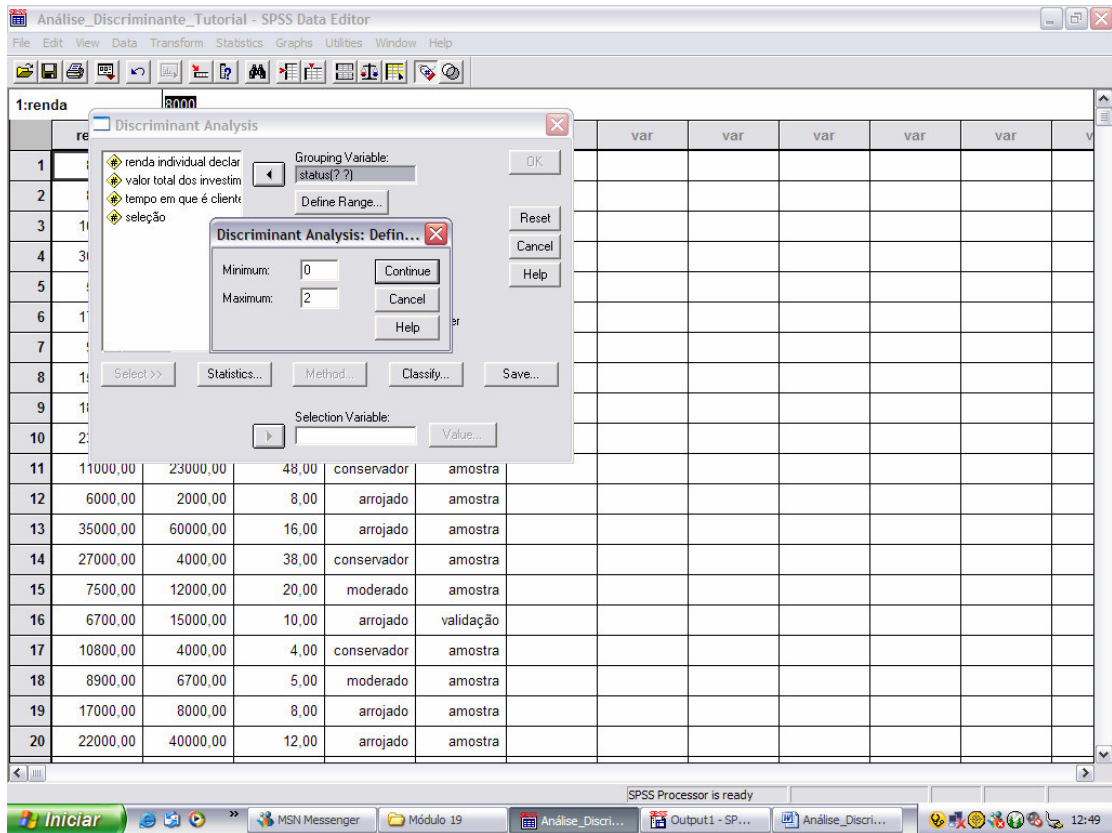


Figura 6: Seleção das categorias da variável dependente.

Ademais à aversão ao risco, o banco considera que o status do investidor é influenciado por outras variáveis (as variáveis independentes). Este estudo incluiu a “renda individual mensal”, “o valor total dos investimentos” e o “tempo em que é cliente do banco”, portanto estas variáveis são selecionadas para “independents”.

Módulo 19 – Análise Discriminante

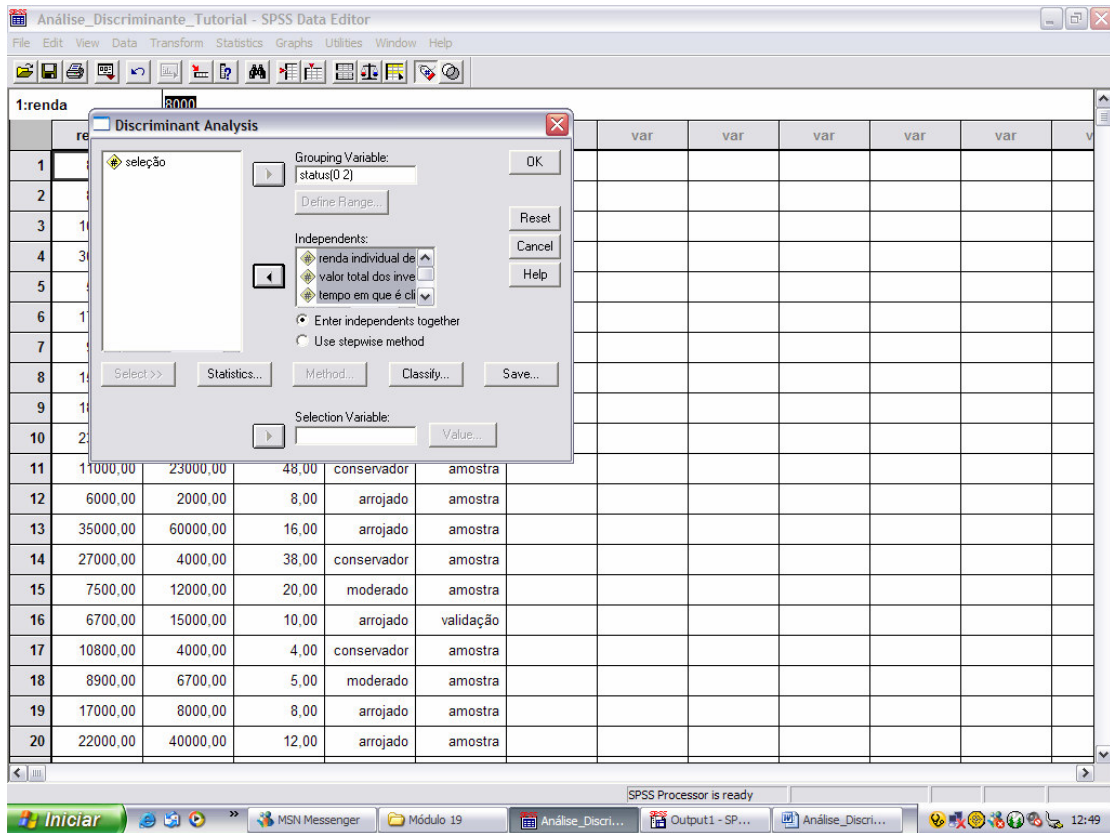


Figura 7: Seleção das variáveis independentes.

Em “select” vamos permitir que se faça a divisão da amostra em amostra de construção e validação. Colocamos a variável “seleção” em “selection variable”.

Módulo 19 – Análise Discriminante

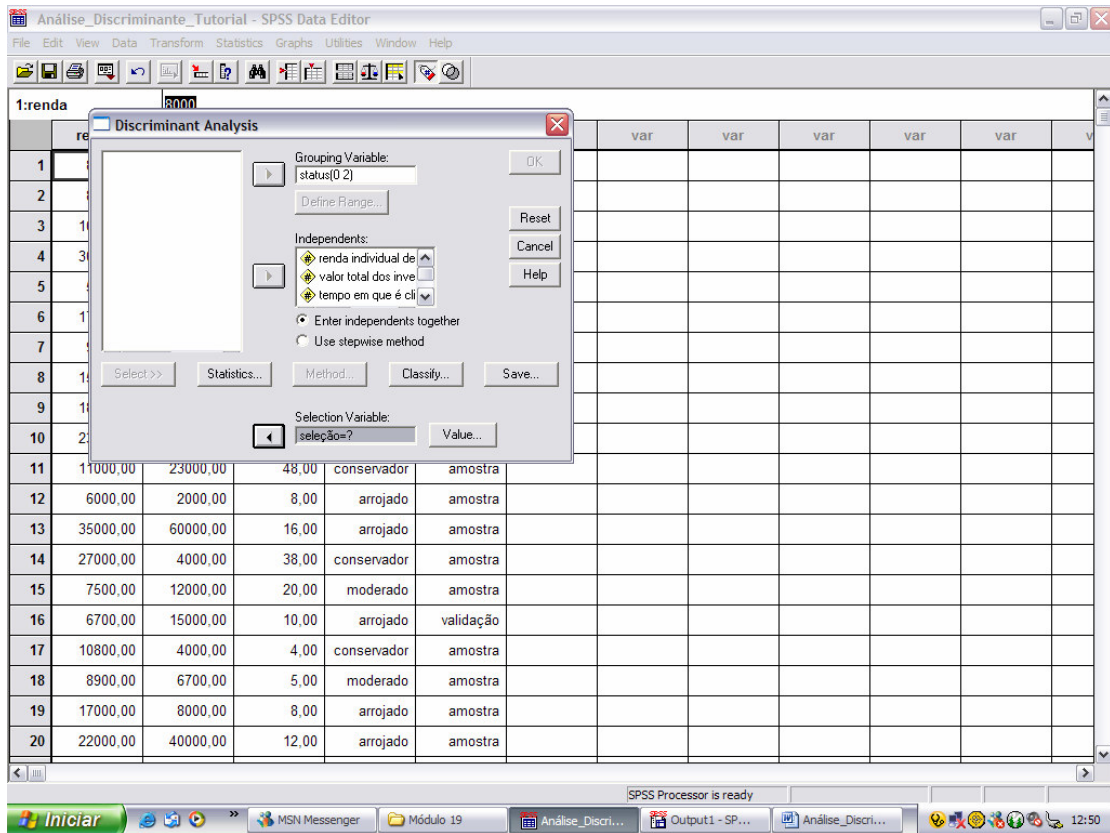


Figura 8: Seleção da variável que determina o grupo de validação.

No botão “value” vamos escolher quais casos servirão para a amostra de construção.

Neste problema a variável “seleção” possui as seguintes categorias:

0= amostra;

1= validação.

Portanto em “value for selection variable” colocaremos o valor 0.

Módulo 19 – Análise Discriminante

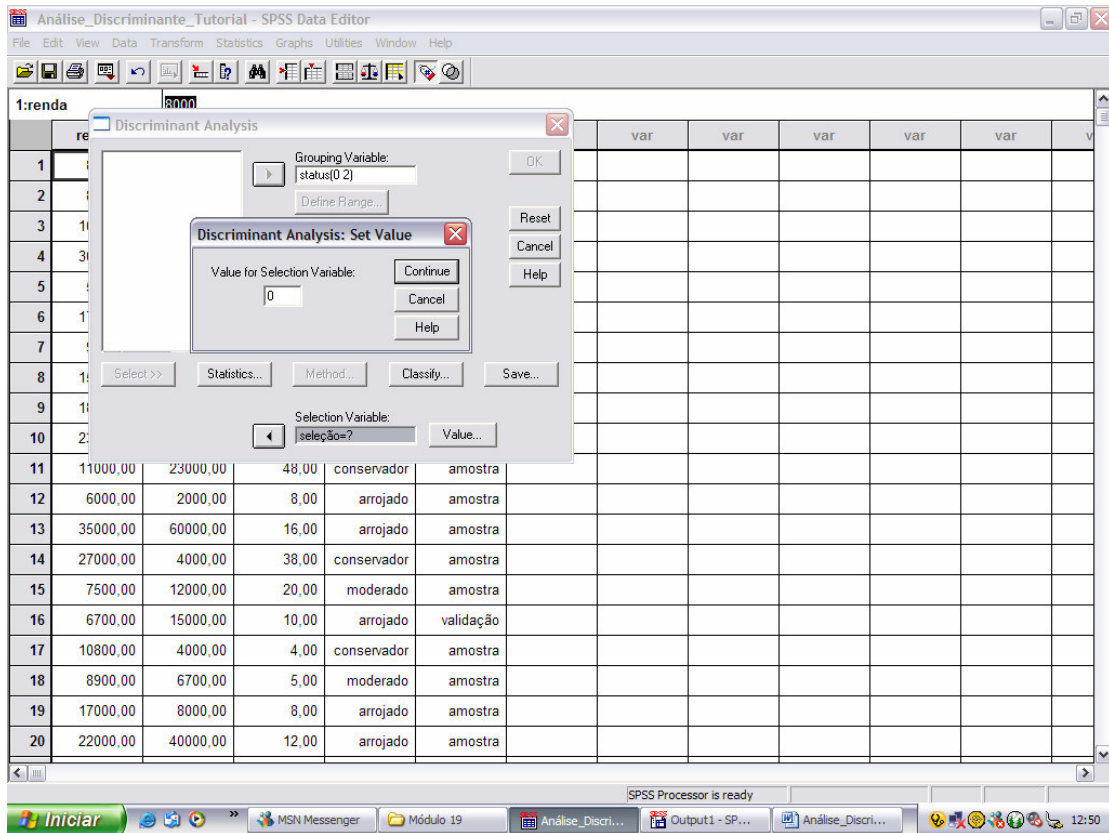


Figura 9: Seleção da categoria da variável de seleção.

No botão “statistics” selecionamos as estatísticas desejadas. Em “descriptives” selecionamos as opções:

“Means”: apresenta as estatísticas descritivas, média e desvio padrão, de cada categoria para cada uma das variáveis independentes (tabela 2);

“Univariate ANOVAs”: apresenta a análise de variância, teste de igualdade de médias entre as três categorias nas variáveis independentes (tabela 3);

“Box’s M”: apresenta o teste M- box, que testa a igualdade das matrizes de covariância (tabela 5 e 6).

Em “function coefficients” podemos pedir ao software coeficientes da análise:

“Fisher’s”: apresenta os coeficientes da função discriminante, considerando-a uma função linear (tabela 15);

“Unstandardized”: apresenta os coeficientes da função discriminante canônica não padronizada (tabela 11).

Módulo 19 – Análise Discriminante

Em “matrizes” podemos pedir matrizes de correlação ou covariância. Seleccionamos somente a opção “Within groups correlation”, que apresenta a correlação das variáveis independentes entre si (tabela 4).

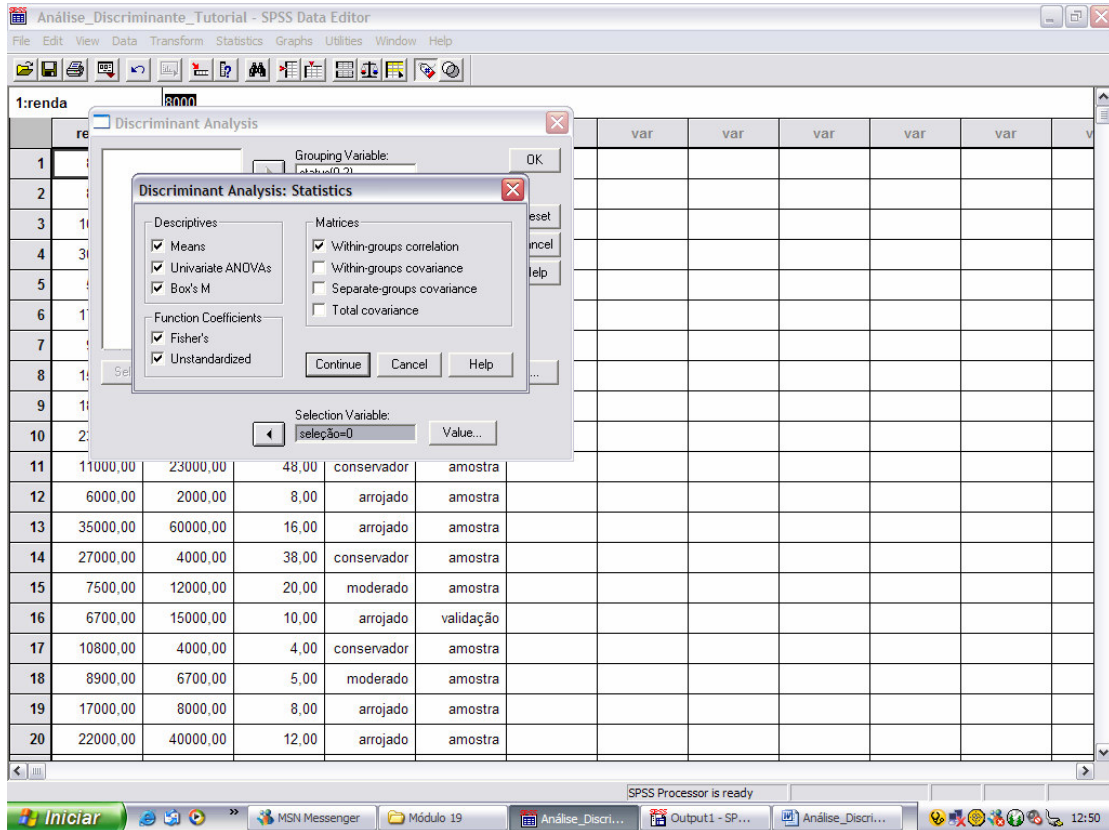


Figura 10: Seleção dos coeficientes desejados.

Em “classification” seleccionamos os resultados desejados sobre a classificação dos casos.

Em “prior probabilities” escolhemos a opção “all groups equal”, que considera as probabilidades de um caso ser classificado em cada um dos grupos igualmente prováveis (tabela 14).

Em “display” seleccionamos:

“Summary table”: apresenta o número total de casos válidos, quantos são utilizados na amostra de construção e de validação (tabelas 1 e 13);

“Leave-one-out classification”: apresenta o resultado final da análise para os casos em que foi realizada a validação cruzada (tabela 16).

Em “use covariance matrix” escolhemos a opção “within groups”.

Módulo 19 – Análise Discriminante

Em “plots” escolhemos os gráficos desejados. Selecionamos a opção “separate- groups” que dará um gráfico de dispersão dos casos da amostra para cada categoria e a opção “territorial map” que apresenta o mapa territorial.

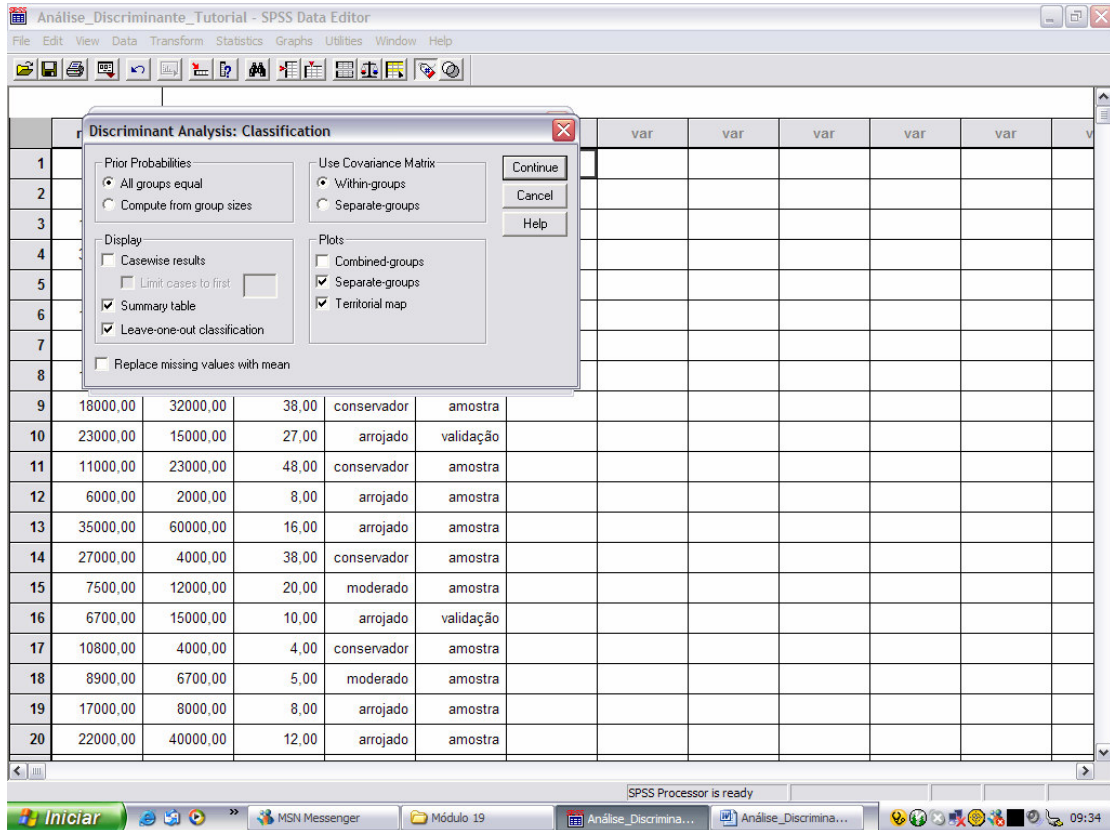


Figura 11: Seleção dos out puts sobre a classificação dos grupos

Em “save” podemos optar por criar ou não uma nova variável. As opções possíveis são:

- “Predicted group membership”: que apresenta a classificação dos casos já existentes de acordo com a função discriminante construída;
- “Discriminant scores”: que apresenta os escores discriminantes para cada caso em cada função discriminante, caso mais de uma função discriminante seja determinada;
- “Probabilities of group membership”: gera variáveis (tantas quantas forem as categorias da variável dependente) que dão a probabilidade do caso estar em cada uma das categorias.

Escolheremos apenas a opção “predicted group membership”.

Módulo 19 – Análise Discriminante

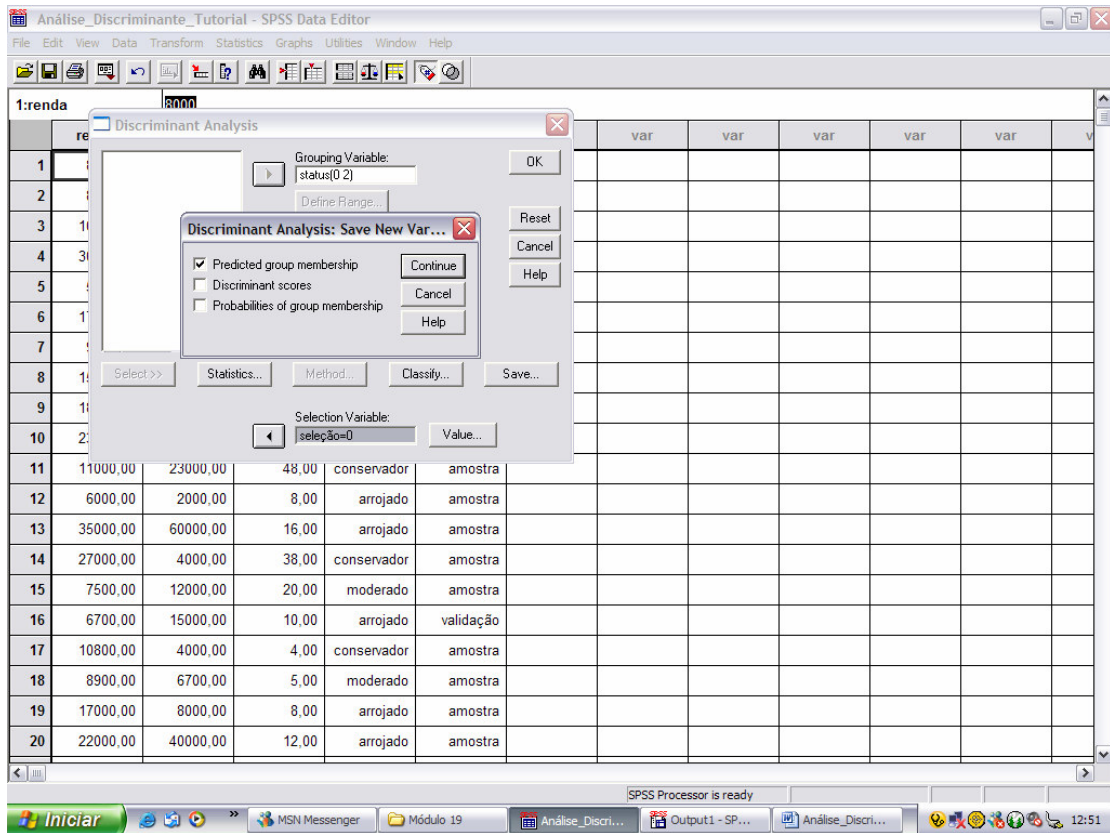
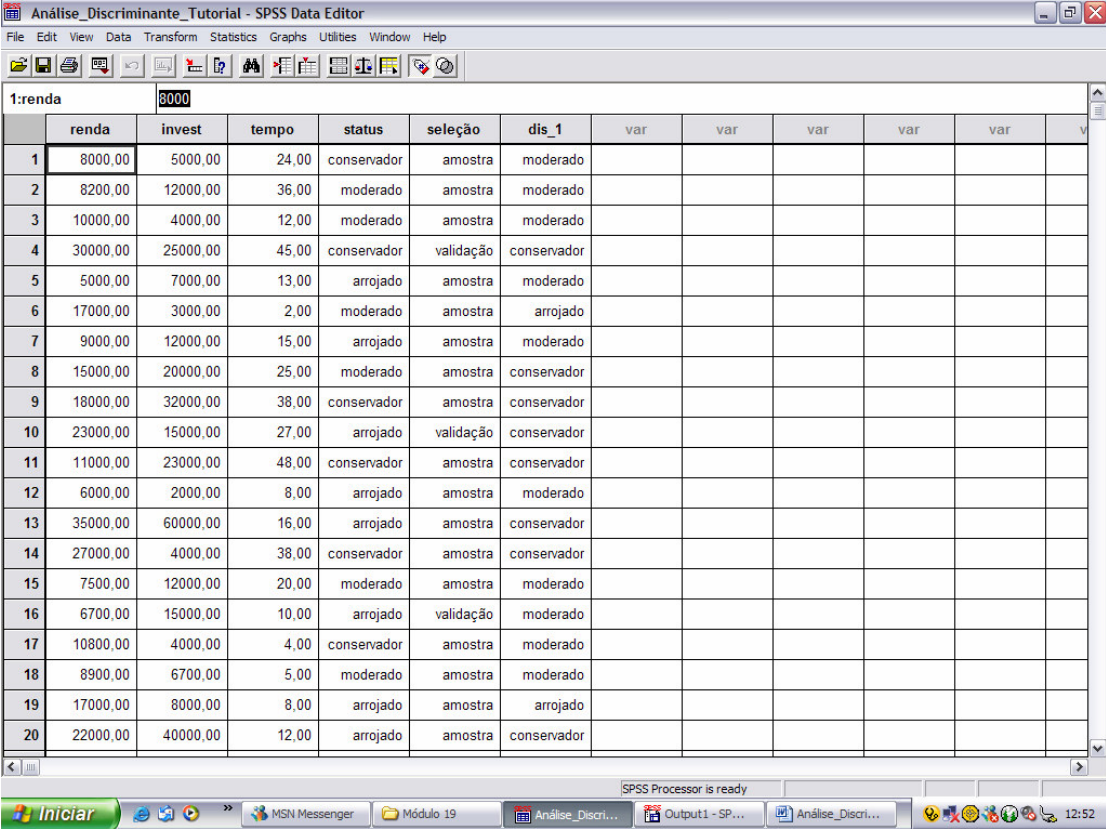


Figura 12: Criação de nova variável.

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

Observe que depois de rodar a análise, a planilha de dados inicial ganha uma nova variável, (solicitada no menu “save”), que apresenta os casos dentro das categorias de investidores, variável “dis_1”.

Módulo 19 – Análise Discriminante



Análise_Discriminante_Tutorial - SPSS Data Editor

File Edit View Data Transform Statistics Graphs Utilities Window Help

1:renda 8000

	renda	invest	tempo	status	seleção	dis_1	var	var	var	var	var	var
1	8000,00	5000,00	24,00	conservador	amostra	moderado						
2	8200,00	12000,00	36,00	moderado	amostra	moderado						
3	10000,00	4000,00	12,00	moderado	amostra	moderado						
4	30000,00	25000,00	45,00	conservador	validação	conservador						
5	5000,00	7000,00	13,00	arrojado	amostra	moderado						
6	17000,00	3000,00	2,00	moderado	amostra	arrojado						
7	9000,00	12000,00	15,00	arrojado	amostra	moderado						
8	15000,00	20000,00	25,00	moderado	amostra	conservador						
9	18000,00	32000,00	38,00	conservador	amostra	conservador						
10	23000,00	15000,00	27,00	arrojado	validação	conservador						
11	11000,00	23000,00	48,00	conservador	amostra	conservador						
12	6000,00	2000,00	8,00	arrojado	amostra	moderado						
13	35000,00	60000,00	16,00	arrojado	amostra	conservador						
14	27000,00	4000,00	38,00	conservador	amostra	conservador						
15	7500,00	12000,00	20,00	moderado	amostra	moderado						
16	6700,00	15000,00	10,00	arrojado	validação	moderado						
17	10800,00	4000,00	4,00	conservador	amostra	moderado						
18	8900,00	6700,00	5,00	moderado	amostra	moderado						
19	17000,00	8000,00	8,00	arrojado	amostra	arrojado						
20	22000,00	40000,00	12,00	arrojado	amostra	conservador						

SPSS Processor is ready

Iniciar MSN Messenger Módulo 19 Análise_Discri... Output1 - SP... Análise_Discri... 12:52

Figura 13: Planilha de dados com a nova variável gerada.