

Como gerar amostras aleatórias simples e estratificadas proporcionais com o Br.Office.org Calc ®

Neste texto mostraremos como usar o Br.Office.org Calc ® para gerar uma amostra aleatória simples e estratificada proporcional. Ao contrário do Microsoft Excel ®, que tem um suplemento estatístico, o Calc pode obter amostras aleatórias apenas de uma maneira: através da função ALEATÓRIOENTRE (). Nosso objetivo é retirar uma amostra aleatória de 250 clientes da montadora Toyord: primeiramente sem considerar a população dividida em estratos, e posteriormente levando em conta eventuais subdivisões. De acordo com o arquivo População Toyord há 6500 clientes disponíveis. Alguns procedimentos adicionais, imprescindíveis, serão apresentados também. A seguir alguns detalhes sobre os dados.

A Megamontadora TOYORD regularmente conduz pesquisas de mercado com os clientes que compraram carros zero km diretamente de suas concessionárias. O objetivo é avaliar a satisfação dos clientes em relação aos diferentes modelos, seu design, adequação ao perfil do cliente. A última pesquisa foi terminada em julho de 2010: 5000 clientes foram entrevistados entre o total de 30000 que compraram veículos novos entre maio de 2009 e maio de 2010. A pesquisa foi restringida aos modelos mais vendidos, e que já estão no mercado há 10 anos. As seguintes variáveis foram obtidas:

- 1) Modelo comprado: o compacto Chiconaultla, o seda médio DeltaForce3, a perua familiar Valentiniana, a van SpaceShuttle ou o esportivo LuxuriousCar.
- 2) Opcionais: inexistentes (apenas os itens de série); ar condicionado e direção hidráulica; ar condicionado, direção hidráulica e trio elétrico; ar condicionado, direção hidráulica, trio elétrico e freios ABS.
- 3) Opinião sobre o design: se os clientes consideram o design do veículo comprado ultrapassado, atualizado, ou adiante dos concorrentes.
- 4) Opinião sobre a concessionária onde comprou o veículo (incluindo atendimento na venda, manutenção programada e eventuais problemas imprevistos): muito insatisfatória, insatisfatória, não causou impressão, satisfatória, bastante satisfatória.
- 5) Opinião geral sobre o veículo adquirido: muito insatisfeito, insatisfeito, satisfeito, bastante satisfeito.
- 6) Renda declarada pelo cliente: em salários mínimos.
- 7) Número de pessoas geralmente transportadas no veículo.
- 8) Quilometragem mensal média percorrida com o veículo.
- 9) Percepção do cliente de há quantos anos o veículo comprado teve a sua última remodelação de design: em anos completos (se há menos de um ano o entrevistador anotou zero).
- 10) Idade do cliente em anos completos.

Apresentaremos os seguintes tópicos: procedimentos para preparação dos dados (necessária para conhecer melhor a população, permitindo identificar eventuais estratos), procedimentos para retirada de uma amostra aleatória simples e procedimentos para retirada de uma amostra estratificada proporcional.

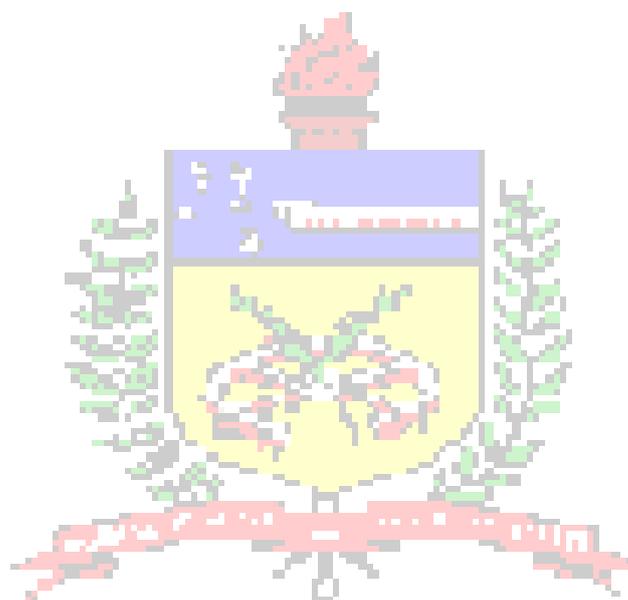
Índice analítico

1. Preparação dos Dados.....	4
1.1 – Filtragem	4
1.1.1 – Filtragem para variáveis qualitativas	4
1.1.2 – Filtragem para variáveis quantitativas	6
1.2 – Recodificação	7
1.3 – Classificação.....	10
2. Amostragem.....	13
2.1 – Amostragem aleatória simples	13
2.2 – Amostragem aleatória estratificada proporcional.....	16
2.2.1 – Tamanho de amostra com erro amostral definido por estrato	19
2.2.2 – Tamanho de amostra com erro amostral definido para a população.....	21
2.2.3 – Obtenção da amostra estratificada proporcional	23

Índice de Figuras

Figura 1 - Menu AutoFiltro	4
Figura 2 - Aplicação de AutoFiltro às variáveis de PopulaçãoToyord	4
Figura 3 - Filtro da variável Modelo	5
Figura 4 - Variável Modelo - Apenas células Vazias	5
Figura 5 - Modelo: registro incorreto Chic Figura 6 - Correção de registro incorreto	5
Figura 7 - Correção de um dos erros de registro de Modelo	6
Figura 8 - Filtro de Renda - Menores Valores.....	6
Figura 9 - Filtro de Renda - Maiores Valores	6
Figura 10 - Dados sobre o cliente com a maior renda	7
Figura 11 - Estrutura de recodificação da variável Renda	8
Figura 12 - Função de recodificação da variável Renda em RendaC	9
Figura 13 - Função de recodificação da variável Idade em IdadeC.....	9
Figura 14 - Variáveis RendaC e IdadeC	9
Figura 15 - Menu Dados - Opção Classificar.....	10
Figura 16 - Critérios de classificação: apenas por RendaC	10
Figura 17 – Opções de classificação.....	11
Figura 18 - Resultados da Classificação em função de RendaC e IdadeC	11
Figura 19 - Modificação dos números dos casos após classificação	12
Figura 20 - Amostragem aleatória simples- início	13
Figura 21 - Função PROC para recuperar dados de Modelo.....	13
Figura 22 - Assistente de Funções do Calc	14
Figura 23 - Assistente de funções: ALEATÓRIOENTRE - 1a parte	14
Figura 24 - Assistente de funções: ALEATÓRIOENTRE - 2a parte	15
Figura 25 - Primeiro elemento da amostra aleatória simples	15
Figura 26 - Amostra aleatória simples de 250 elementos - parcial.....	15
Figura 27 - Acréscimo de colunas com número dos casos.....	16
Figura 28 - Caso com células vazias	16
Figura 29 - Início do estrato 1 a 4 s.m. - 18 a 25 anos Figura 30 - Final do estrato 1 a 4 s.m. - 18 a 25 anos	17
Figura 31 - Final dos estratos 1 a 4 s.m. - 25 a 40 anos e 1 a 4 s.m. - mais de 40 anos	17
Figura 32 - Final dos estratos 4 a 12 s.m. - 18 a 25 anos e 4 a 12 s.m. – 25 a 40 anos	17
Figura 33 - Final do estrato 4 a 12 s.m. – mais de 40 anos e células vazias.....	18
Figura 34 - Final do estrato Mais de 12 s.m. – 18 a 25 anos	18
Figura 35 - Final dos estratos Mais de 12 s.m. - 25 a 40 anos e Mais de 12 s.m. – Mais de 40 anos	18
Figura 36 - Estratos em função de RendaC e IdadeC	19
Figura 37 - Estratos em função de RendaC e IdadeC – sem Mais de 12 s.m. – 18 a 25 anos	19
Figura 38 – Cálculo de n_0 para o estrato 1 a 4 s.m. – 18 a 25 anos com $E_0 = 2,5\%$	20
Figura 39 – Cálculo de n para o estrato 1 a 4 s.m. – 18 a 25 anos com $E_0 = 2,5\%$	20
Figura 40 – Função ARREDONDAR.PARA.CIMA	20
Figura 41 – Função ARREDONDAR.PARA.CIMA - Opções	21
Figura 42 – Tamanhos de amostra para os estratos com $E_0 = 2,5\%$	21
Figura 43 – Cálculo de n_0 para $E_0 = 2,5\%$ (toda a população)	22
Figura 44 – Cálculo de n para $E_0 = 2,5\%$ (toda a população).....	22
Figura 45 - Arredondamento para cima do valor de n	22
Figura 46 - Arredondamento para cima do valor de n no primeiro estrato	22
Figura 47 – Tamanhos de amostra para cada estrato	23
Figura 48 – Colunas Ordem e Sorteado.....	23
Figura 49 – Sorteio do primeiro elemento do primeiro estrato	23
Figura 50 - Resultados da amostra do primeiro estrato - final	24

Figura 51 – Sorteio do primeiro elemento do segundo estrato.....24
Figura 52 – Resumo dos estratos (casos na população) e ordem de sorteio (na amostra).....24



**Marcelo
Menezes
Reis**

1. Preparação dos Dados

A preparação dos dados é indispensável para que possamos identificar as características da população que serão importantes para o processo de amostragem. Entre todos os procedimentos disponíveis vamos estudar Filtragem de variáveis, Recodificação de variáveis e Classificação do arquivo de dados em função de uma ou mais variáveis

1.1 – Filtragem

A filtragem é uma ferramenta extremamente útil para a análise de dados. Permite realizar a pré-análise dos dados, realizar buscas específicas por uma informação e corrigir eventuais erros. Vamos apresentar os procedimentos no Calc para executar filtragem de variáveis qualitativas e quantitativas, permitindo a identificação de valores perdidos e erros de registro (e sua correção), valores discrepantes, entre outras informações.

Abra o arquivo PopulaçãoToyord.ods e procure pela planilha “Dados”. Vamos trabalhar com duas variáveis: Modelo (qualitativa) e Renda (quantitativa contínua).

1.1.1 – Filtragem para variáveis qualitativas

Para acionar o filtro automático do Calc, para todas as variáveis, basta pôr o cursor em qualquer célula ocupada da planilha “Dados”. Depois, no menu “Dados” procure por “Filtro” e depois pela opção “AutoFiltro” tal como na Figura 1.

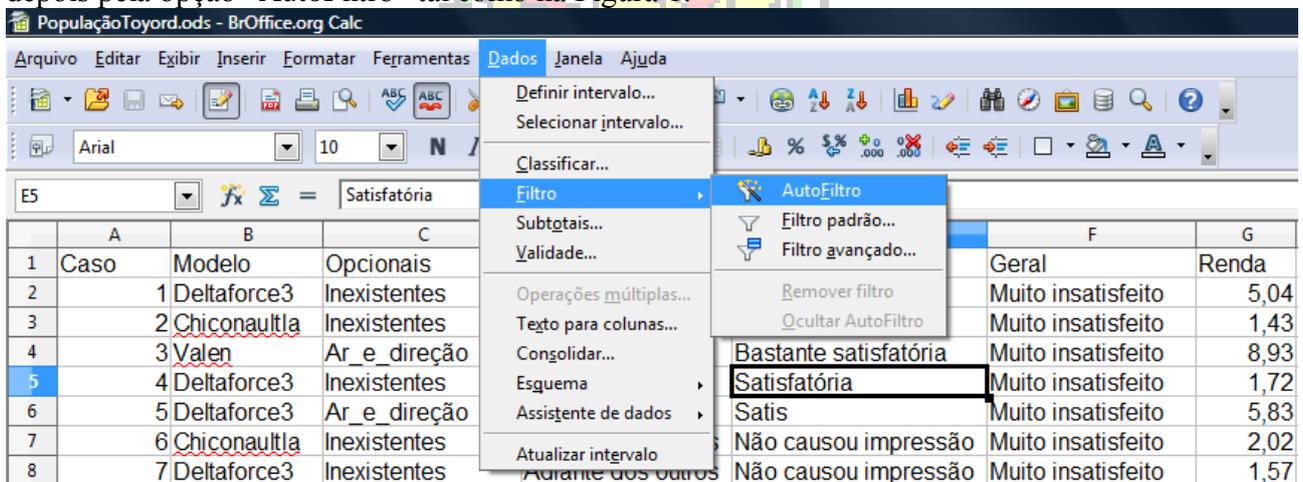


Figura 1 - Menu AutoFiltro

Ao escolher AutoFiltro o Calc automaticamente aplica os botões de filtro a todas as variáveis, e o resultado está na Figura 2.

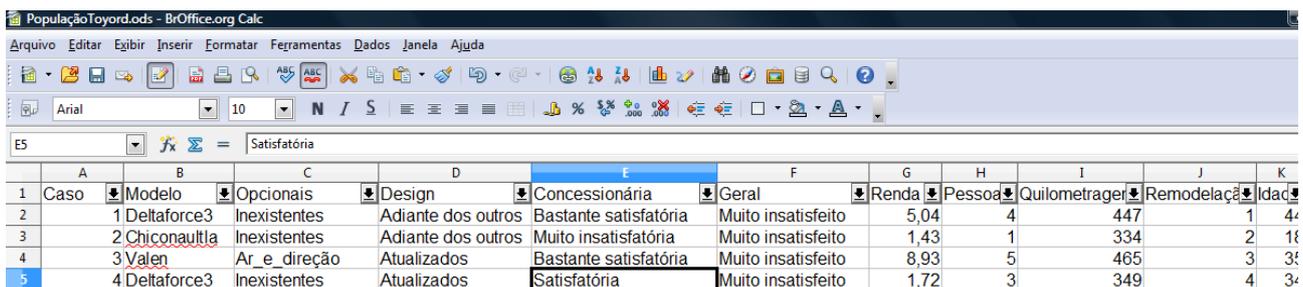


Figura 2 - Aplicação de AutoFiltro às variáveis de PopulaçãoToyord

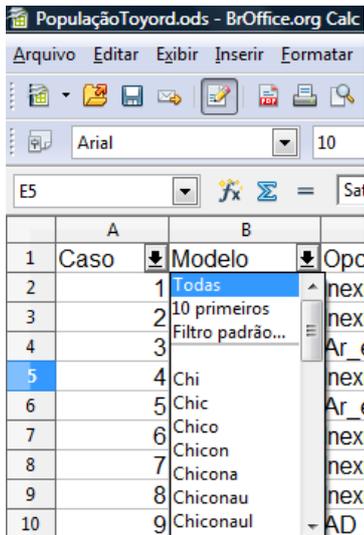


Figura 3 - Filtro da variável Modelo

Os valores corretos para a variável Modelo são Chiconaultla, Deltaforce3, Valentiniana, SpaceShuttle e LuxuriousCar. Mas, há vários registros incorretos para todos, em que houve truncamento de letras. É possível ver também uma linha em branco, que são as células vazias, se selecionada permitirá avaliar a quantidade de dados perdidos; “10 primeiros”; “Filtro padrão”, na qual podemos construir o filtro de nosso interesse.

Selecionando qualquer dos valores o Calc apresentará apenas as células que o contém. Selecionando a linha em branco obtemos a Figura 4.

	A	B
1	Caso	Modelo
181	180	
264	263	
485	484	
806	805	
812	811	
1301	1300	
1317	1316	
2303	2302	
2417	2416	

Figura 4 - Variável Modelo - Apenas células Vazias

Os dados perdidos encontram-se nas células B181, B264, B806, B812, B1301, B1317, B2303 e B2417, totalizando 8 observações.

Como são apenas 7 em 5000 registros, representando 0,16%, bem abaixo de 5%, podemos considerar uma quantidade aceitável. Porém, se possível, devemos investigar as causas deste “sumiço de dados”, e se possível, descobrir seu conteúdo (deve haver algum registro do cliente 180 (célula B181) em algum lugar, que diga qual o modelo escolhido).

Para corrigir os erros basta selecionar um dos valores incorretamente registrados, como “Chic” na Figura 3. Os resultados podem ser vistos na Figura 5

	A	B	C
1	Caso	Modelo	Opç
951	950	Chic	Ar_
962	961	Chic	Inex
2048	2047	Chic	Inex

Figura 5 - Modelo: registro incorreto Chic

	A	B	C
1	Caso	Modelo	Opç
951	950	Chiconaultla	
962	961	Chic	
2048	2047	Chic	

Figura 6 - Correção de registro incorreto

O erro Chic (deveria ser Chiconaultla) aparece nas células B951, B962 e B2048. Para corrigi-los precisamos digitar corretamente o valor: na Figura 6 corrige-se Chic para Chiconaultla. MUITO IMPORTANTE: no Calc a correção dos erros precisa ser feita *INDIVIDUALMENTE*, não é possível arrastar” o valor corrigido para todas as células incorretas¹. Se fizéssemos isso na Figura 6 todas as células entre B951 e B2048 seriam transformadas em Chiconaultla, o que não é desejado.

¹ O Microsoft Excel permite fazer isso sem modificar as outras células do intervalo.

Precisamos então corrigir todos os erros de registro, para todas as variáveis quantitativas, manualmente, como foi feito com Chic, resultando na Figura 7.

	A	B	C
1	Caso	Modelo	Opcionais
951	950	Chiconaultla	Ar_e_direção
962	961	Chiconaultla	Inexistentes
2048	2047	Chiconaultla	Inexistentes

Figura 7 - Correção de um dos erros de registro de Modelo

Os dados sem erros estão no arquivo PopulaçãoToyordSemErros.ods, que será usado de agora em diante.

1.1.2 – Filtragem para variáveis quantitativas

Podemos usar o mesmo procedimento das variáveis qualitativas nas quantitativas. Se realizarmos o procedimento de filtragem com a variável Renda, quantitativa contínua, vamos observar a grande quantidade de valores possíveis entre o mínimo (1 salário mínimo), mostrado na Figura 8, e o máximo (95,73 salários mínimos), mostrado na Figura 9.

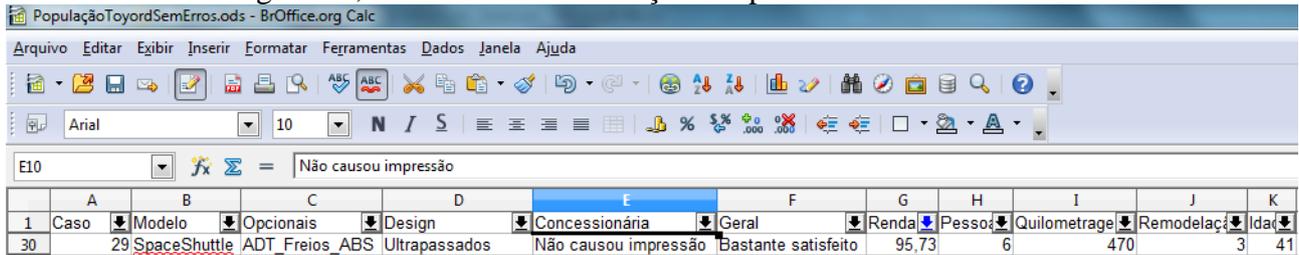
	A	B	C	D	E	F	G	H
1	Caso	Modelo	Opcionais	Design	Concessionária	Geral	Renda	Pessoa
2	1	Deltaforce3	Inexistentes	Adiante dos outros	Bastante satisfatória	Muito insatisfeito	Todas	
3	2	Chiconaultla	Inexistentes	Adiante dos outros	Muito insatisfatória	Muito insatisfeito	10 primeiros	
4	3	Valentiniana	Ar_e_direção	Atualizados	Bastante satisfatória	Muito insatisfeito	Filtro padrão...	
5	4	Deltaforce3	Inexistentes	Atualizados	Satisfatória	Muito insatisfeito	1	
6	5	Deltaforce3	Ar_e_direção	Atualizados	Satisfatória	Muito insatisfeito	1,005	
7	6	Chiconaultla	Inexistentes	Adiante dos outros	Não causou impressão	Muito insatisfeito	1,01	
8	7	Deltaforce3	Inexistentes	Adiante dos outros	Não causou impressão	Muito insatisfeito	1,015	
9	8	Deltaforce3	Inexistentes	Atualizados	Muito insatisfatória	Insatisfeito	1,02	
10	9	Deltaforce3	AD_Trio_Elétrico	Atualizados	Não causou impressão	Insatisfeito	1,025	
11	10	Valentiniana	Inexistentes	Atualizados	Satisfatória	Muito insatisfeito	1,03	
12	11	Valentiniana	Ar_e_direção	Atualizados	Não causou impressão	Muito insatisfeito	1,035	

Figura 8 - Filtro de Renda - Menores Valores

	A	B	C	D	E	F	G	H
1	Caso	Modelo	Opcionais	Design	Concessionária	Geral	Renda	Pessoa
2	1	Deltaforce3	Inexistentes	Adiante dos outros	Bastante satisfatória	Muito insatisfeito	46,69	
3	2	Chiconaultla	Inexistentes	Adiante dos outros	Muito insatisfatória	Muito insatisfeito	47,36	
4	3	Valentiniana	Ar_e_direção	Atualizados	Bastante satisfatória	Muito insatisfeito	48,04	
5	4	Deltaforce3	Inexistentes	Atualizados	Satisfatória	Muito insatisfeito	48,3	
6	5	Deltaforce3	Ar_e_direção	Atualizados	Satisfatória	Muito insatisfeito	49,355	
7	6	Chiconaultla	Inexistentes	Adiante dos outros	Não causou impressão	Muito insatisfeito	60,72	
8	7	Deltaforce3	Inexistentes	Adiante dos outros	Não causou impressão	Muito insatisfeito	61,26	
9	8	Deltaforce3	Inexistentes	Atualizados	Muito insatisfatória	Insatisfeito	62,26	
10	9	Deltaforce3	AD_Trio_Elétrico	Atualizados	Não causou impressão	Insatisfeito	69,7	
11	10	Valentiniana	Inexistentes	Atualizados	Satisfatória	Muito insatisfeito	95,73	
12	11	Valentiniana	Ar_e_direção	Atualizados	Não causou impressão	Muito insatisfeito		

Figura 9 - Filtro de Renda - Maiores Valores

Se quisermos saber mais sobre o cliente com a maior renda basta selecionar o valor 95,73 e teremos a tela da Figura 10, com todas as informações disponíveis.



	A	B	C	D	E	F	G	H	I	J	K
1	Caso	Modelo	Opcionais	Design	Concessionária	Geral	Renda	Pessoas	Quilometragem	Remodelação	Idade
30	29	SpaceShuttle	ADT_Freios_ABS	Ultrapassados	Não causou impressão	Bastante satisfeito	95,73	6	470	3	41

Figura 10 - Dados sobre o cliente com a maior renda

Trata-se de um cliente que optou pelo modelo SpaceShuttle, com todos os opcionais (ar condicionado, direção hidráulica, trio elétrico e freios ABS), que considera os modelos da Toyord ultrapassados, a concessionária não causou impressão nele, mas no geral está bastante satisfeito. Ele costuma rodar com 6 pessoas a bordo (o que provavelmente explica a escolha pela van), roda em média 470 km, acha que os veículos da Toyord foram remodelados há 3 anos, e tem 41 anos completos de idade.

Poderíamos repetir o procedimento para qualquer outro valor de Renda.

1.2 – Recodificação

Em muitas situações de análise de dados pode haver interesse em criar novas variáveis a partir das existentes: com a finalidade de agrupar valores de uma variável qualitativa ou quantitativa, ou transformar uma variável quantitativa em qualitativa. Com isso torna-se possível realizar novas análises dos dados, sem modificar os dados originais.

Podemos fazer isso para variáveis qualitativas ou quantitativas. Para as qualitativas usualmente busca-se reduzir o número de opções. Por exemplo, poderíamos agrupar as opiniões “Bastante satisfeito” e “Satisfeito” na variável Geral e recodificá-la em uma nova variável como “Positiva”, e as demais em “Negativa”.

A recodificação de variáveis quantitativas pode ter duas finalidades:

- transformá-la em outra variável quantitativa (quando se chama transformação, mediante alguma operação matemática);
- transformá-la em uma variável qualitativa (quando se chama realmente recodificação, mediante a aplicação de operações lógicas, tal como as vistas para variáveis qualitativas).

Em ambos os casos o objetivo é obter uma nova variável que facilite a análise dos dados, tornando-a mais resumida ou significativa. Por exemplo, vamos realizar uma recodificação das variáveis Renda e Idade, no arquivo PopulaçãoToyord:

- recodificar Renda em uma variável qualitativa, criando faixas de valores que definirão clientes de renda de 1 a 4 salários mínimos, 4 a 12 salários mínimos e de mais de 12 salários mínimos.
- recodificar Idade em uma variável qualitativa, criando faixas de valores que definirão clientes de idade de 18 a 25 anos, 25 a 40 anos e de mais de 40 anos.

Neste momento é importante uma pequena introdução sobre as funções lógicas do Calc. Há várias disponíveis, sendo as principais, com a sua sintaxe:

- SE(teste lógico; ação caso o teste lógico resulte verdadeiro; ação caso o teste lógico resulte falso). É possível aninhar até sete funções SE como argumentos valor_se_verdadeiro e valor_se_falso para construir testes mais elaborados.

- E(teste lógico1; teste lógico2; ...). Retornará VERDADEIRO se todos os testes lógicos resultarem verdadeiros; retornará FALSO se um ou mais testes lógicos resultarem falsos. Teste lógico1; teste lógico2;... são de 1 a 30 condições para testar e que podem ser VERDADEIRO ou FALSO.
- OU(teste lógico1; teste lógico2; ...). Retorna VERDADEIRO se pelo menos um dos testes lógicos resultar verdadeiro; retorna FALSO se todos os testes lógicos resultarem FALSOS. Teste lógico1; teste lógico2,... são de uma a 30 condições que você deseja testar e que podem resultar em VERDADEIRO ou FALSO.

Ao fazer recodificação ou transformação de variáveis, é preciso tomar cuidado com as células vazias, pois vamos “criar” informações ao codificá-la em qualitativa, ou causar erro ao transformá-la.

Podemos agora pensar na recodificação de Renda em uma variável qualitativa, que chamaremos RendaC, que ocupará a coluna L. Vamos criar três classes *arbitrárias* de renda: os clientes com Renda de até 4 salários mínimos serão chamados RendaC “1 a 4 s.m.”, os com Renda entre 4 e 12 serão chamados RendaC “4 a 12 s.m.”, e os com Renda acima de 12 salários mínimos serão RendaC “Mais de 12 s.m.”. Trata-se de um caso que exige a utilização de funções SE: precisaremos de duas para a classificação da Renda, e mais uma para lidar com as células vazias. Veja a estrutura na Figura 11.

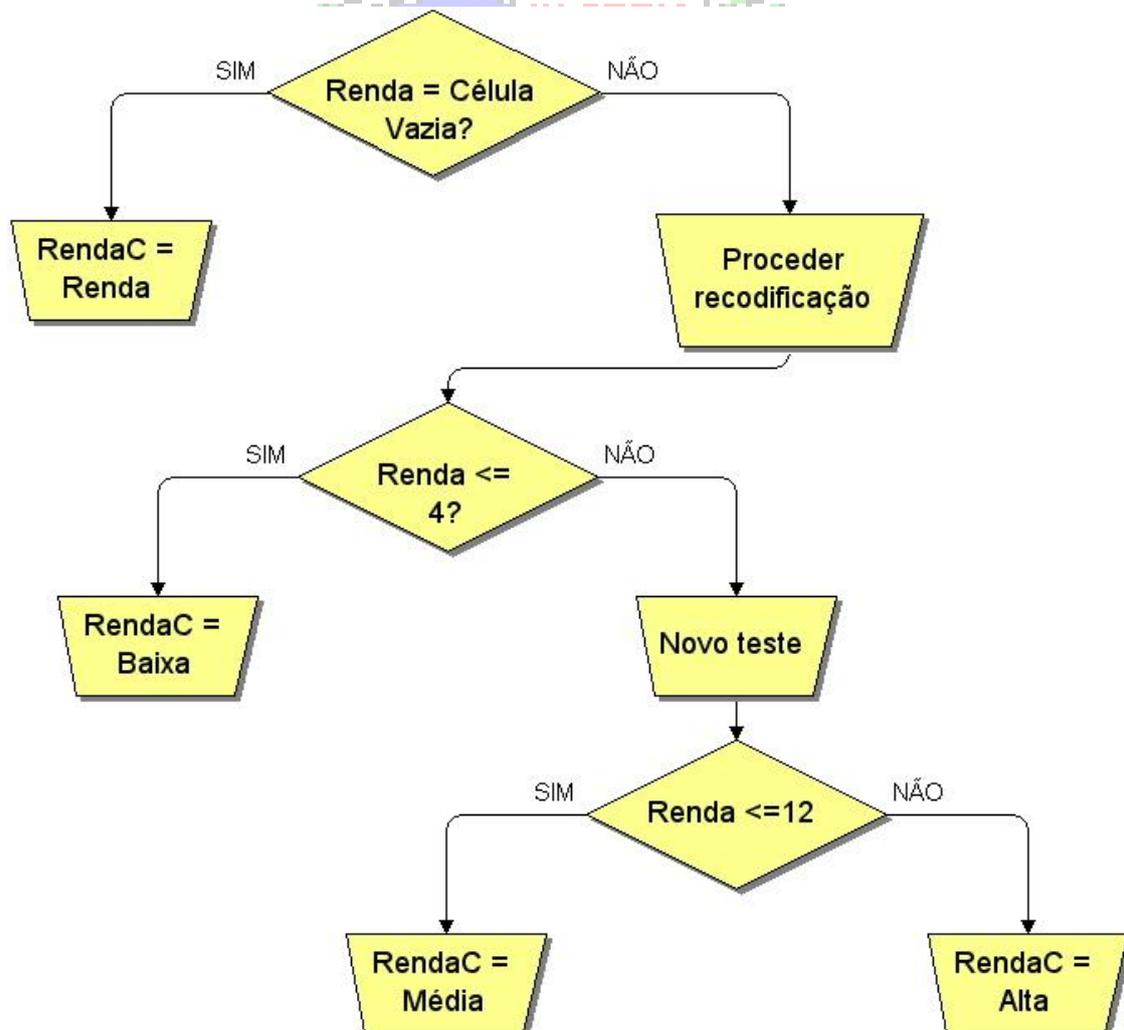


Figura 11 - Estrutura de recodificação da variável Renda

Vamos ver os passos:

- 1) Se a célula de Renda for vazia a célula de RendaC também será, para evitar a criação de informação.
- 2) Se a célula de Renda NÃO for vazia pode-se proceder fazer a recodificação propriamente dita.
- 3) Se a célula de Renda for menor ou igual a 4, a célula de RendaC será igual à “1 a (cliente com renda baixa).
- 4) Se a célula de Renda NÃO for menor ou igual a 4, faz-se novo teste.
- 5) Se a célula de Renda for menor ou igual a 12 (já se sabe pelo teste anterior que é maior do que 4 salários mínimos), a célula de RendaC será igual à Média (cliente com renda média).
- 6) Se a célula de Renda NÃO for menor ou igual a 12, já que o teste anterior verificou que é maior do 4 salários mínimos, só resta a possibilidade de ser maior do que 12. Então a célula de RendaC será igual à Alta (cliente com renda alta).

Observe a implementação da recodificação da variável Renda no Calc (Figura 29), na planilha “Dados” do arquivo PopulaçãoToyordSemErros.ods.

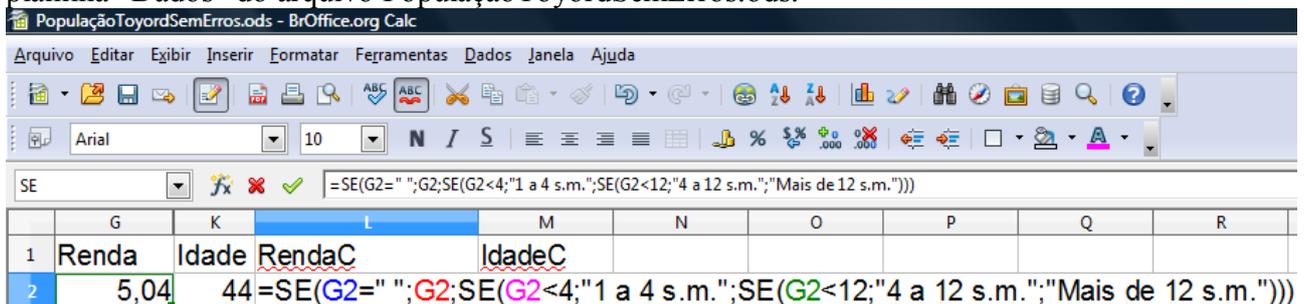


Figura 12 - Função de recodificação da variável Renda em RendaC

Ao arrastar a fórmula até a célula L5001 completamos a recodificação da variável, cujos resultados podem ser vistos na Figura 14.

Podemos recodificar a variável Idade também, de forma análoga vamos criar três categorias: 18 a 25 anos, 25 a 40 anos e acima de 40 anos. A implementação está mostrada na Figura 13, e os resultados também podem ser vistos na Figura 14.

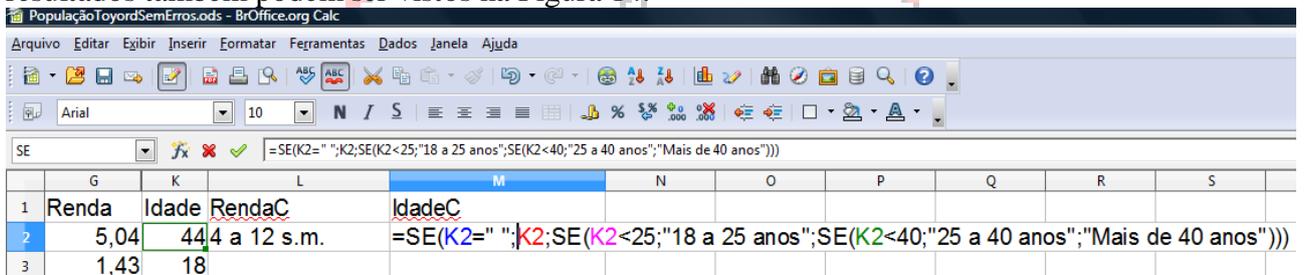


Figura 13 - Função de recodificação da variável Idade em IdadeC

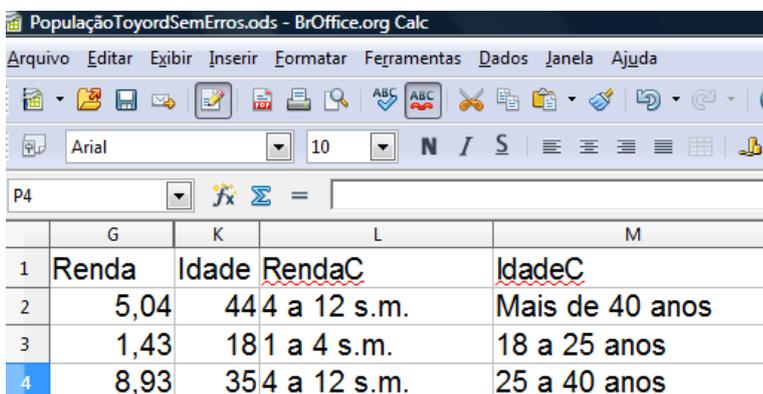


Figura 14 - Variáveis RendaC e IdadeC

Agora há duas novas variáveis qualitativas, relacionadas às variáveis quantitativas de origem: as informações originais não foram perdidas, e temos mais uma maneira de caracterizar o conjunto de dados.

1.3 – Classificação

Em muitos casos há interesse em reordenar o conjunto de dados de maneira a facilitar a visualização dos valores específicos de uma ou mais variáveis. Isso será extremamente útil nas próximas aulas quando construirmos tabelas para os dados. O Calc dispõe de vários mecanismos automáticos de classificação, que podem ser acessados pelo menu Dados, opção Classificar, como na Figura 15, na planilha “Dados” do arquivo PopulaçãoToyordSemErros.ods.

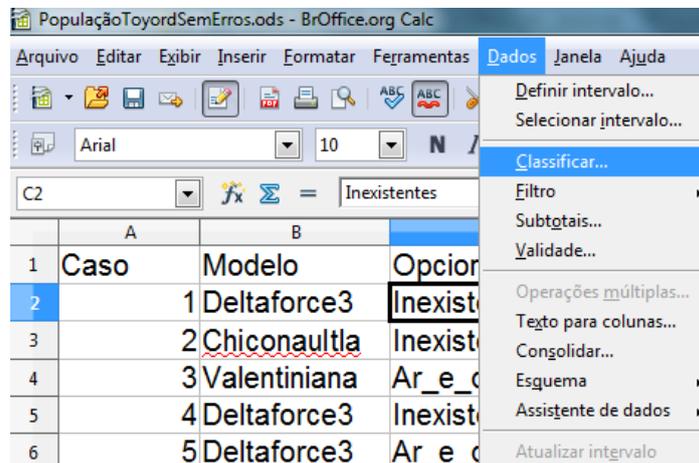


Figura 15 - Menu Dados - Opção Classificar

Podemos classificar os dados em função de até 3 variáveis, de forma automática. Se desejássemos fazer a classificação apenas em função de RendaC e depois por IdadeC bastaria colocá-las na primeira e segunda opções da tela da Figura 16.

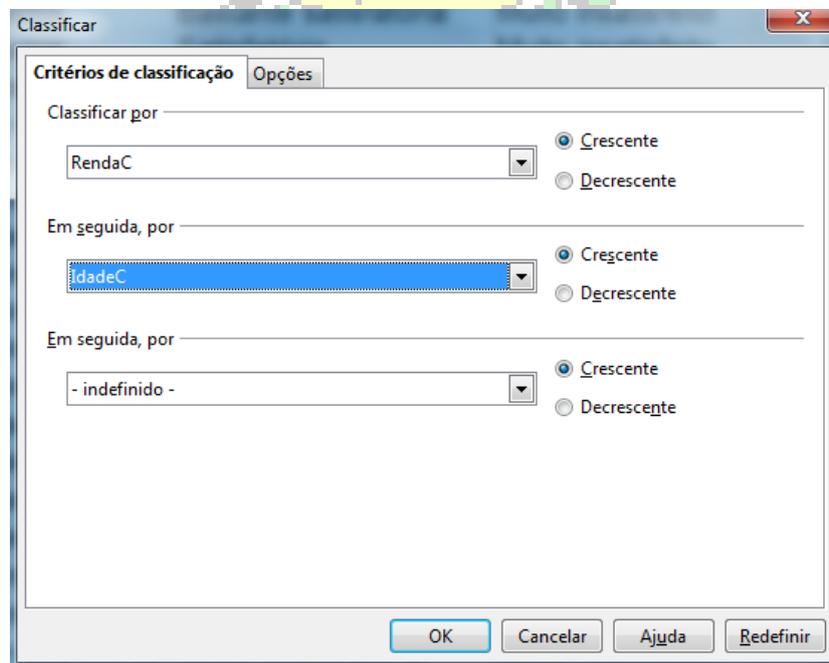


Figura 16 - Critérios de classificação: apenas por RendaC

Ao escolher a aba “Opções” podemos detalhar melhor o processo de classificação, o que resulta na Figura 17.

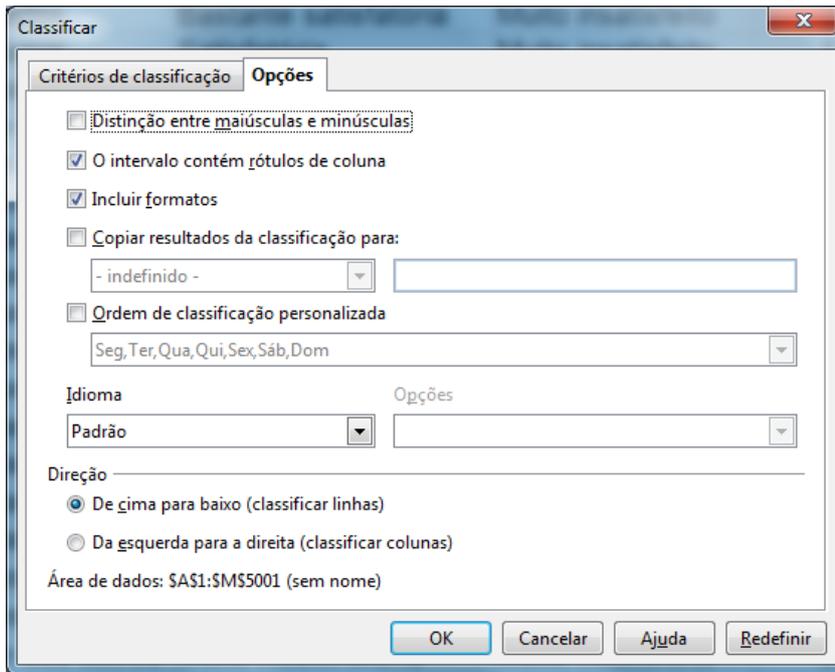


Figura 17 – Opções de classificação

Pressionando OK na Figura 17 os dados são classificados, e o resultado pode ser visto na Figura 18, já salvo como arquivo PopulaçãoToyordSemErrosClassificado.odt

Na tela imediatamente abaixo vê-se que os números dos casos foram reordenados de acordo com os resultados de RendaC e IdadeC. Na tela ao lado estão os valores de RendaC: o mais “baixo” é a célula vazia, depois 1 a 4 s.m., etc. Observe que há valores de IdadeC para as células vazias de RendaC, e eles estão ordenados (18 a 25 anos, 25 a 40 anos, Mais de 40 anos). Todos os outros valores estão ordenados em função de RendaC e IdadeC. Vamos modificar a numeração dos casos, para que a amostragem depois possa ser melhor realizada – Figura 19.

O intervalo de A1 a M5001 contém todas as variáveis, incluindo a linha 1, com os nomes, e as recodificadas também. Podemos fazer distinções entre maiúsculas e minúsculas, copiar os resultados de classificação para um novo intervalo de células (preservando o conjunto de dados original). Muito importante: no nosso caso queremos que a direção da classificação seja “De cima para baixo”, para que as linhas sejam classificadas.

	A	B	C
1	Caso	Modelo	Opcionais
2	3103	Chiconaultla	Inexistentes
3	891	Valentiniana	ADT_Freios_ABS
4	2217	SpaceShuttle	Ar_e_direção
5	1564	Valentiniana	AD_Trio_Elétrico

	L	M
1	RendaC	IdadeC
2		18 a 25 anos
3		25 a 40 anos
4		Mais de 40 anos
5		Mais de 40 anos
6	1 a 4 s.m.	
7	1 a 4 s.m.	
8	1 a 4 s.m.	
9	1 a 4 s.m.	
10	1 a 4 s.m.	
11	1 a 4 s.m.	
12	1 a 4 s.m.	
13	1 a 4 s.m.	
14	1 a 4 s.m.	
15	1 a 4 s.m.	
16	1 a 4 s.m.	18 a 25 anos
17	1 a 4 s.m.	18 a 25 anos

Figura 18 - Resultados da Classificação em função de RendaC e IdadeC

Basta substituir os dois primeiros números dos casos (ver Figura 18) por 1 e 2, como mostrado na tela à esquerda na Figura 19, e depois arrastá-los até a célula A5001. O resultado pode ser visto na tela à direita à Figura 19.

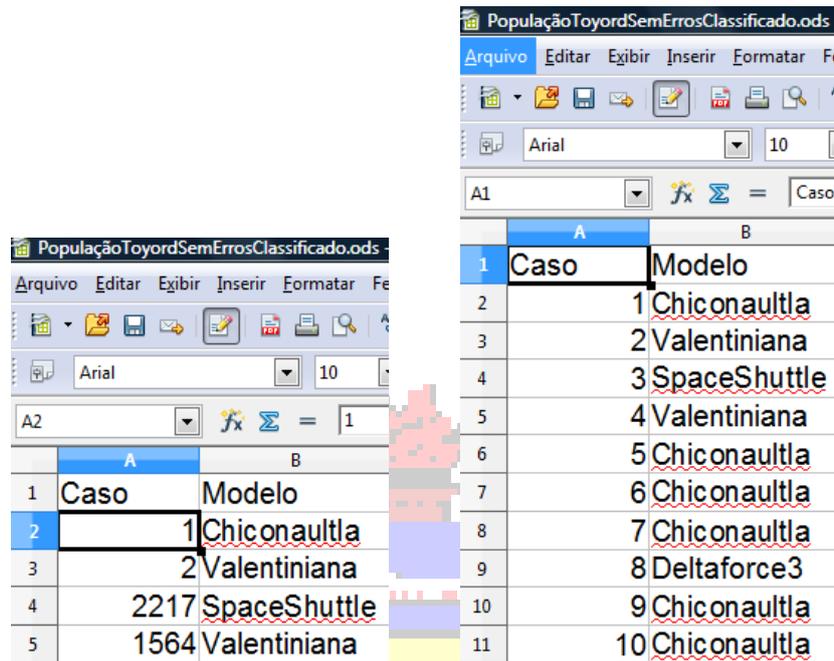


Figura 19 - Modificação dos números dos casos após classificação

Agora os dados estão classificados em função de RendaC e IdadeC: na prática temos 9 estratos possíveis:

RendaC	IdadeC
1 a 4 s.m.	18 a 25 anos
1 a 4 s.m.	25 a 40 anos
1 a 4 s.m.	Mais de 40 anos
4 a 12 s.m.	18 a 25 anos
4 a 12 s.m.	25 a 40 anos
4 a 12 s.m.	Mais de 40 anos
Mais de 12 s.m.	18 a 25 anos
Mais de 12 s.m.	25 a 40 anos
Mais de 12 s.m.	Mais de 40 anos

Se imaginarmos que as duas variáveis influenciam na escolha do modelo de automóvel, ao realizar uma pesquisa por amostragem precisamos considerar os nove estratos formados: elementos dos nove estratos precisam fazer parte da amostra, na mesma proporção encontrada na população, o que configura uma amostra estratificada proporcional². Portanto, precisamos identificar os casos na população, o que será possível devido à reordenação dos números mostrada na Figura 19.

Mas, antes de passarmos à amostragem estratificada proporcional, vamos ver como implementar a amostragem aleatória simples.

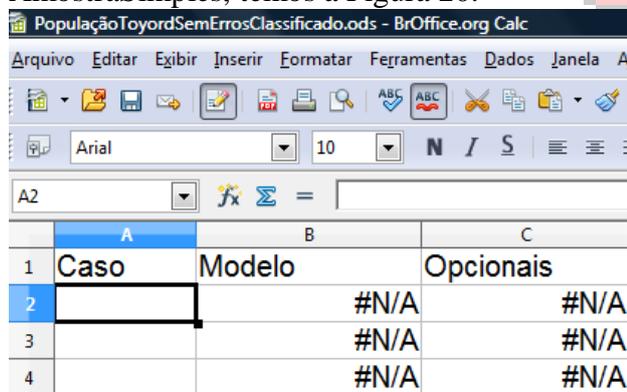
² O tipo de amostragem que permite a melhor representatividade da população.

2. Amostragem

Agora que já temos conhecimento sobre as principais características da população podemos passar a obtenção das amostras propriamente ditas: por meio de amostragem aleatória simples e amostragem aleatória estratificada proporcional.

2.1 – Amostragem aleatória simples

Se for possível considerar a população homogênea o procedimento é simples. Basta definir os números mínimo e máximo dos casos na população e usar a função ALEATÓRIO.ENTRE para selecionar a amostra, no intervalo de casos da população e obter a amostra. Vamos usar o arquivo PopulaçãoToyordSemErrosClassificado.ods, que tem os estratos definidos em função das variáveis RendaC e IdadeC (recodificações de Renda e Idade, respectivamente). Escolhendo a planilha AmostraSimples, temos a Figura 20.

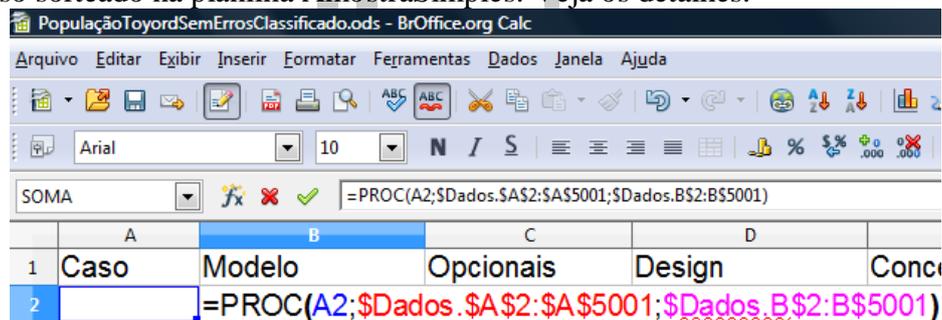


	A	B	C
1	Caso	Modelo	Opcionais
2		#N/A	#N/A
3		#N/A	#N/A
4		#N/A	#N/A

A coluna A contém a variável Caso, que registrará os casos sorteados dentre aqueles disponíveis na planilha Dados. Observe que nas células à direita há os símbolos #N/A. Trata-se de mensagens de erro, há uma *função* que recupera as informações dos casos da planilha Dados a partir dos casos sorteados. Como as células da coluna A ainda estão vazias as mensagens de erro aparecem. A função em questão é a PROC, que descrevemos a seguir.

Figura 20 - Amostragem aleatória simples- início

Na Figura 21 é possível avaliar a função PROC que recupera o valor de Modelo (na planilha Dados) a partir do caso sorteado na planilha AmostraSimples. Veja os detalhes:



	A	B	C	D	E
1	Caso	Modelo	Opcionais	Design	Conce
2		=PROC(A2;\$Dados.\$A\$2:\$A\$5001;\$Dados.B\$2:B\$5001)			

Figura 21 - Função PROC para recuperar dados de Modelo

- A2 é a célula que tem o caso que será sorteado;
- Dados é a planilha de origem, onde estão os casos da população, o intervalo \$A\$2:\$A\$5001 contém todos os casos da população (usamos o \$ antes das letras das colunas e números das linhas para manter as referências absolutas);
- o intervalo \$B\$2:\$B\$5001 (também na planilha Dados) contém os valores de Modelo (novamente usamos o \$ para manter as referências absolutas);
- de acordo com o caso sorteado, a função varrerá os 5000 casos da população (no intervalo A2 a A5001 na planilha Dados) e recuperará o valor de Modelo correspondente a ele (que está no intervalo B2 a B5001, também na planilha Dados);
- para que PROC funcione *o primeiro intervalo precisa estar em ordem crescente* (o que é nosso caso, as células A2 a A5001 contêm os casos 1 a 5000);
- podemos arrastar a função até o fim da amostra, por exemplo, se a amostra tivesse 250 elementos,

até a célula B251 (por isso é importante o uso de referências absolutas);
 - para as outras variáveis basta arrastar a célula B2 até a K2, e substituir o intervalo $\$B\$2:\$B\5001 pelos correspondentes de cada variável ($\$C\$2:\$C\5001 , $\$D\$2:\$D\5001 , etc.).

Imagine que queremos uma amostra de 250 elementos. Basta selecionar a célula A2, tal como mostrado na Figura 21, e selecionar o Assistente de funções, que resulta na tela da Figura 22.

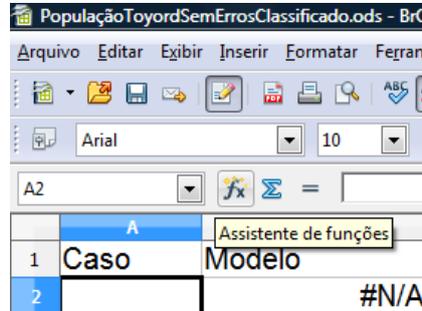


Figura 22 - Assistente de Funções do Calc

Ao pressionar o botão do Assistente de Funções surge a tela da Figura 23.

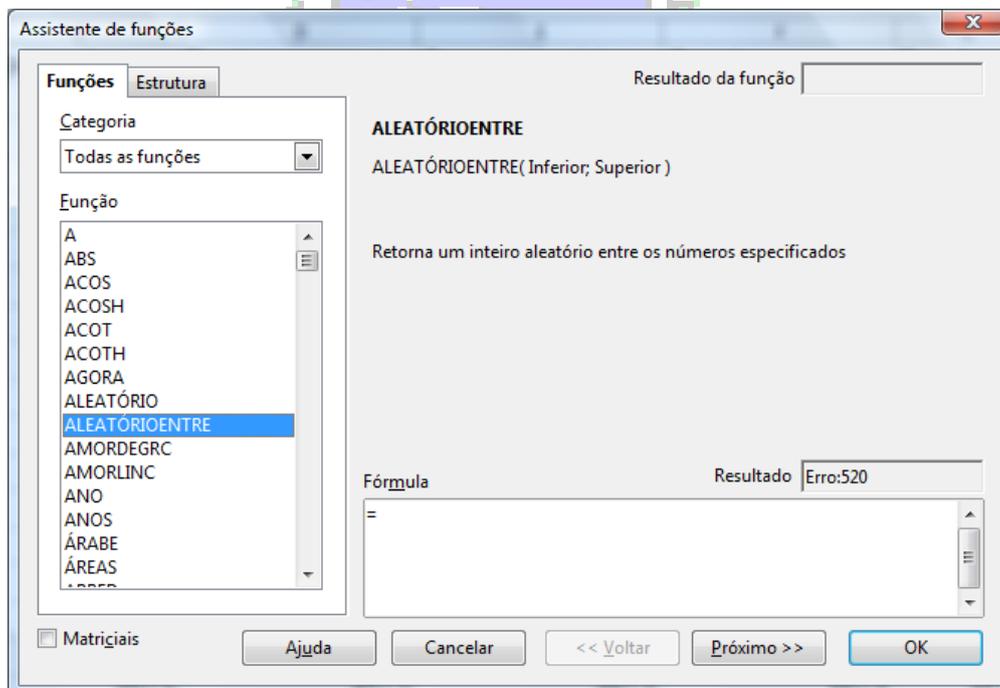
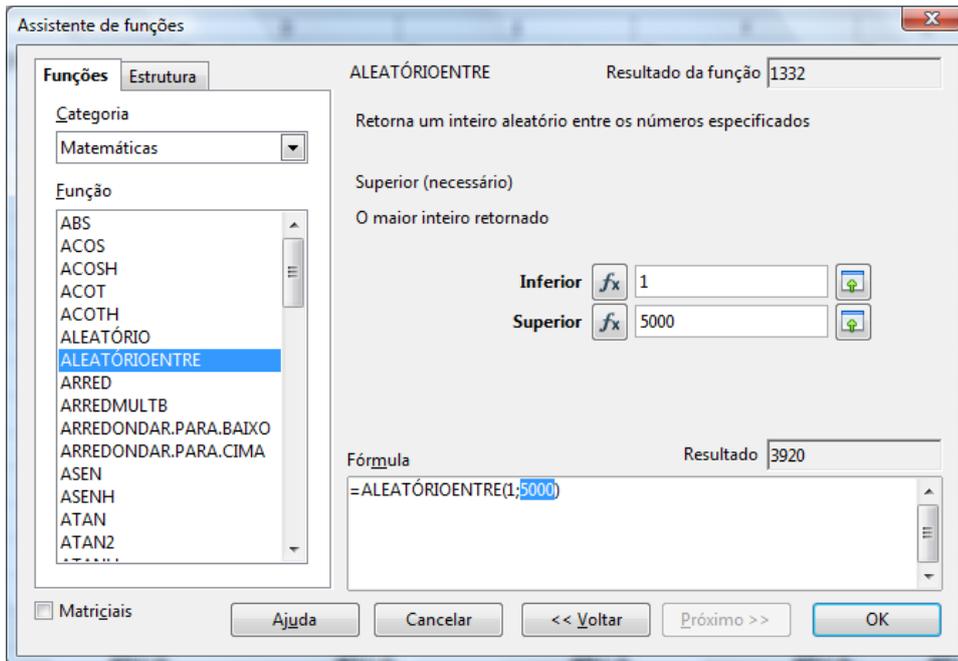


Figura 23 - Assistente de funções: ALEATÓRIOENTRE - 1a parte

Há uma infinidade de funções disponíveis. É possível escolher por categorias específicas (opção “Categoria”), ou selecionar do conjunto de Todas as funções. Como queremos uma amostra aleatória podemos escolher a função ALEATÓRIOENTRE, que retorna um inteiro aleatório entre os números especificados. Como queremos uma amostra aleatória simples, desconsiderando eventuais estratos, basta pressionar “Próximo” na Figura 23 e chegar à Figura 24.



Em “Inferior” é posto o número do primeiro caso (1) e em “Superior” o último (5000). Pressionando “OK” seleciona-se o primeiro integrante da amostra, visto na Figura 25, o caso 1673. As funções PROC obtêm as informações de todas as variáveis para o caso 1673: o cliente adquiriu um modelo Chiconaultla, sem opcionais, etc.

Figura 24 - Assistente de funções: ALEATÓRIOENTRE - 2a parte³

	A	B	C
1	Caso	Modelo	Opcionais
2	1673	Chiconaultla	Inexistentes
3	#N/A	#N/A	#N/A
4	#N/A	#N/A	#N/A

Se arrastarmos a célula A2 até a linha 251 vamos obter a amostra aleatória simples de 250 elementos que desejávamos, da qual uma parte é mostrada na Figura 26.

Figura 25 - Primeiro elemento da amostra aleatória simples

	A	B
1	Caso	Modelo
2	1673	Chiconaultla
3	2733	Valentiniana
4	3852	Valentiniana
5	399	Chiconaultla
6	146	Deltaforce3
7	2940	Chiconaultla
8	349	Chiconaultla
9	2575	LuxuriousCar
10	4567	LuxuriousCar
11	3992	Valentiniana
12	2775	Chiconaultla

Figura 26 - Amostra aleatória simples de 250 elementos - parcial

³ IMPORTANTE: quando você for retirar a amostra o Calc provavelmente selecionará um valor diferente, já que se trata de um processo aleatório.

2.2 – Amostragem aleatória estratificada proporcional

Se houver razão para suspeitar que subdivisões na população possam influenciar nos resultados da pesquisa torna-se recomendável levá-las em conta na amostragem, de maneira a garantir a representatividade da amostra. No caso da pesquisa da Toyord, supomos que as variáveis Renda e Idade influenciam na escolha do modelo de automóvel, então precisamos considerá-las na amostragem. Nas seções 1.2 e 1.3 mostramos como recodificar as variáveis e como classificar o arquivo de dados em função dos seus valores, resultando em nove estratos. Os elementos dos nove estratos precisam fazer parte da amostra, na mesma proporção encontrada na população, o que configura uma amostra estratificada proporcional⁴. Portanto, precisamos identificar os casos na população, o que será possível devido à reordenação dos números mostrada na Figura 19 (ver seção 1.2). Para facilitar a identificação podemos acrescentar mais uma coluna com os números dos casos à direita das colunas das variáveis recodificadas, como mostrado na Figura 27.

	L	M	N
1	RendaC	IdadeC	Caso
2		18 a 25 anos	1
3		25 a 40 anos	2
4		Mais de 40 anos	
5		Mais de 40 anos	
6	1 a 4 s.m.		
7	1 a 4 s.m.		
8	1 a 4 s.m.		

Basta selecionar as células N2 e N3 e arrastá-las até a linha 5001, obtendo toda a sequência. Na Figura 28 é possível observar que há um total de 14 casos com células vazias de RendaC (casos 1 a 4) ou de IdadeC (casos 5 a 14). Como não temos como recuperar as informações referentes a tais casos, e devido ao fato que representam apenas 14 observações dentre 5000 (0,28%), podemos desconsiderá-las, e começar a amostragem a partir do caso 15, com os clientes de renda de 1 a 4 s.m. e idade de 18 a 25 anos.

Figura 27 - Acréscimo de colunas com número dos casos

	L	M	N
1	RendaC	IdadeC	Caso
2		18 a 25 anos	1
3		25 a 40 anos	2
4		Mais de 40 anos	3
5		Mais de 40 anos	4
6	1 a 4 s.m.		5
7	1 a 4 s.m.		6
8	1 a 4 s.m.		7
9	1 a 4 s.m.		8
10	1 a 4 s.m.		9
11	1 a 4 s.m.		10
12	1 a 4 s.m.		11
13	1 a 4 s.m.		12
14	1 a 4 s.m.		13
15	1 a 4 s.m.		14
16	1 a 4 s.m.	18 a 25 anos	15
17	1 a 4 s.m.	18 a 25 anos	16

Figura 28 - Caso com células vazias

⁴ Outras variáveis poderiam ser consideradas na estratificação, teoricamente não há limite para o número de variáveis estratificadoras, mas, é claro, há limites práticos para a segmentação.

Temos que identificar os casos de cada estrato, que serão depois usados na função ALEATÓRIOENTRE. Se selecionarmos as células N15 e N16 (Figura 29) e formos “descendo” na planilha vamos chegar até a linha 375, que indica que os casos do estrato 1 a 4 s.m. – 18 a 25 anos estão entre 15 e 374, e que o estrato seguinte (1 a 4 s.m. – 25 a 40 anos) começa no caso 375, como mostrado na Figura 30.

	L	M	N
15	1 a 4 s.m.		14
16	1 a 4 s.m.	18 a 25 anos	15
17	1 a 4 s.m.	18 a 25 anos	16
18	1 a 4 s.m.	18 a 25 anos	17
19	1 a 4 s.m.	18 a 25 anos	18

Figura 29 - Início do estrato 1 a 4 s.m. - 18 a 25 anos

	L	M	N
374	1 a 4 s.m.	18 a 25 anos	373
375	1 a 4 s.m.	18 a 25 anos	374
376	1 a 4 s.m.	25 a 40 anos	375
377	1 a 4 s.m.	25 a 40 anos	376

Figura 30 - Final do estrato 1 a 4 s.m. - 18 a 25 anos

Então o primeiro estrato vai do caso 15 ao 374, o segundo do 375 ao 2053, e o terceiro do 2054 ao 2406 (ver Figura 31)

	L	M	N
2053	1 a 4 s.m.	25 a 40 anos	2052
2054	1 a 4 s.m.	25 a 40 anos	2053
2055	1 a 4 s.m.	Mais de 40 anos	2054
2056	1 a 4 s.m.	Mais de 40 anos	2055

	L	M	N
2406	1 a 4 s.m.	Mais de 40 anos	2405
2407	1 a 4 s.m.	Mais de 40 anos	2406
2408	4 a 12 s.m.		2407
2409	4 a 12 s.m.		2408
2410	4 a 12 s.m.		2409
2411	4 a 12 s.m.		2410
2412	4 a 12 s.m.		2411
2413	4 a 12 s.m.	18 a 25 anos	2412
2414	4 a 12 s.m.	18 a 25 anos	2413

Figura 31 - Final dos estratos 1 a 4 s.m. - 25 a 40 anos e 1 a 4 s.m. - mais de 40 anos

Observe que após o estrato 1 a 4 s.m. – Mais de 40 anos aparecem mais células vazias, que também deverão ser desconsideradas. O próximo estrato (4 a 12 s.m. – 18 a 25 anos) começará no caso 2412. Podemos ver a evolução nas Figuras 32 e 33: quarto estrato de 2412 a 2486, quinto de 2487 a 3572 (Figura 32), sexto de 3573 a 4205, com mais algumas células vazias (Figura 33).

	L	M	N
2486	4 a 12 s.m.	18 a 25 anos	2485
2487	4 a 12 s.m.	18 a 25 anos	2486
2488	4 a 12 s.m.	25 a 40 anos	2487
2489	4 a 12 s.m.	25 a 40 anos	2488

	L	M	N
3572	4 a 12 s.m.	25 a 40 anos	3571
3573	4 a 12 s.m.	25 a 40 anos	3572
3574	4 a 12 s.m.	Mais de 40 anos	3573
3575	4 a 12 s.m.	Mais de 40 anos	3574

Figura 32 - Final dos estratos 4 a 12 s.m. - 18 a 25 anos e 4 a 12 s.m. - 25 a 40 anos

	L	M	N
4205	4 a 12 s.m.	Mais de 40 anos	4204
4206	4 a 12 s.m.	Mais de 40 anos	4205
4207	Mais de 12 s.m.		4206
4208	Mais de 12 s.m.		4207
4209	Mais de 12 s.m.		4208
4210	Mais de 12 s.m.	18 a 25 anos	4209
4211	Mais de 12 s.m.	18 a 25 anos	4210

Figura 33 - Final do estrato 4 a 12 s.m. – mais de 40 anos e células vazias

Mais uma vez, desconsideraremos as células vazias. Observe que o estrato Mais de 12 s.m. – 18 a 25 anos só tem dois casos (4209 e 4210), na Figura 34, e podemos ver a evolução dos outros na Figura 35: oitavo estrato de 4211 a 4551, e nono de 4552 a 5000.

	L	M	N
4210	Mais de 12 s.m.	18 a 25 anos	4209
4211	Mais de 12 s.m.	18 a 25 anos	4210
4212	Mais de 12 s.m.	25 a 40 anos	4211
4213	Mais de 12 s.m.	25 a 40 anos	4212

Figura 34 - Final do estrato Mais de 12 s.m. – 18 a 25 anos

	L	M	N
4551	Mais de 12 s.m.	25 a 40 anos	4550
4552	Mais de 12 s.m.	25 a 40 anos	4551
4553	Mais de 12 s.m.	Mais de 40 anos	4552
4554	Mais de 12 s.m.	Mais de 40 anos	4553

	L	M	N
4991	Mais de 12 s.m.	Mais de 40 anos	4990
4992	Mais de 12 s.m.	Mais de 40 anos	4991
4993	Mais de 12 s.m.	Mais de 40 anos	4992
4994	Mais de 12 s.m.	Mais de 40 anos	4993
4995	Mais de 12 s.m.	Mais de 40 anos	4994
4996	Mais de 12 s.m.	Mais de 40 anos	4995
4997	Mais de 12 s.m.	Mais de 40 anos	4996
4998	Mais de 12 s.m.	Mais de 40 anos	4997
4999	Mais de 12 s.m.	Mais de 40 anos	4998
5000	Mais de 12 s.m.	Mais de 40 anos	4999
5001	Mais de 12 s.m.	Mais de 40 anos	5000

Figura 35 - Final dos estratos Mais de 12 s.m. - 25 a 40 anos e Mais de 12 s.m. – Mais de 40 anos

Um resumo dos estratos, com números dos casos, e percentual em relação ao total de células não vazias, pode ser visto na Figura 36.

	P	Q	R	S	T	U
4	ESTRATOS				N	%
5	1 a 4 s.m.	18 a 25 anos	15	374	360	7,23%
6	1 a 4 s.m.	25 a 40 anos	375	2053	1679	33,73%
7	1 a 4 s.m.	Mais de 40 anos	2054	2406	353	7,09%
8	4 a 12 s.m.	18 a 25 anos	2412	2486	75	1,51%
9	4 a 12 s.m.	25 a 40 anos	2487	3572	1086	21,82%
10	4 a 12 s.m.	Mais de 40 anos	3573	4205	633	12,72%
11	Mais de 12 s.m.	18 a 25 anos	4209	4210	2	0,04%
12	Mais de 12 s.m.	25 a 40 anos	4211	4551	341	6,85%
13	Mais de 12 s.m.	Mais de 40 anos	4552	5000	449	9,02%
14					4978	100,00%

Figura 36 - Estratos em função de RendaC e IdadeC

Observe que o estrato Mais de 12 s.m. – 18 a 25 anos representa apenas 0,04% da população de 4978 clientes, o que permite desconsiderá-lo na amostragem, resultando na tabela da Figura 37.

	X	Y	Z	AA	AB	AC
4	ESTRATOS				N	%
5	1 a 4 s.m.	18 a 25 anos	15	374	360	7,23%
6	1 a 4 s.m.	25 a 40 anos	375	2053	1679	33,74%
7	1 a 4 s.m.	Mais de 40 anos	2054	2406	353	7,09%
8	4 a 12 s.m.	18 a 25 anos	2412	2486	75	1,51%
9	4 a 12 s.m.	25 a 40 anos	2487	3572	1086	21,82%
10	4 a 12 s.m.	Mais de 40 anos	3573	4205	633	12,72%
11	Mais de 12 s.m.	25 a 40 anos	4211	4551	341	6,85%
12	Mais de 12 s.m.	Mais de 40 anos	4552	5000	449	9,02%
13					4976	100,00%

Figura 37 - Estratos em função de RendaC e IdadeC – sem Mais de 12 s.m. – 18 a 25 anos

O estrato 1 a 4 s.m. – 25 a 40 anos é o mais numeroso, portanto deve ter o maior número de elementos na amostra. Já o estrato 4 a 12 s.m. – 18 a 25 anos, representa apenas 1,51% da população, devendo ter a menor representação proporcional na amostra.

Agora podemos realizar a amostragem. Vamos usar dois procedimentos de cálculo do tamanho de amostra: calculando usando o erro amostral *dentro* de cada estrato, e desconsiderando os estratos inicialmente e depois dividindo a amostra de forma proporcional.

2.2.1 – Tamanho de amostra com erro amostral definido por estrato

Neste caso queremos que o erro amostral dentro de cada estrato seja o mesmo. Então precisamos realizar os cálculos do tamanho de amostra por estrato: obtém-se a primeira estimativa n_0 , corrige-se esta estimativa usando como tamanho de população N o tamanho de cada estrato.

Para um erro amostral de 2,5% (0,025) podemos ver a fórmula de n_0 para o primeiro estrato na Figura 38: basta arrastar a fórmula da célula AD5 a AD12. Os seus resultados (1600 para todos os estratos) e a fórmula de n para o primeiro estrato (usando n_0 e o tamanho do estrato) estão na Figura 39. Arrastando a fórmula da célula AE5 a AE12 temos todos os tamanhos de amostra, por estrato.

	X	Y	Z	AA	AB	AC	AD	AE
4	ESTRATOS				N	%	e0	n0
5	1 a 4 s.m.	18 a 25 anos	15	374	360	7,23%	0,025	=1/(AD5^2)
6	1 a 4 s.m.	25 a 40 anos	375	2053	1679	33,74%	0,025	
7	1 a 4 s.m.	Mais de 40 anos	2054	2406	353	7,09%	0,025	
8	4 a 12 s.m.	18 a 25 anos	2412	2486	75	1,51%	0,025	
9	4 a 12 s.m.	25 a 40 anos	2487	3572	1086	21,82%	0,025	
10	4 a 12 s.m.	Mais de 40 anos	3573	4205	633	12,72%	0,025	
11	Mais de 12 s.m.	25 a 40 anos	4211	4551	341	6,85%	0,025	
12	Mais de 12 s.m.	Mais de 40 anos	4552	5000	449	9,02%	0,025	
13					4976	100,00%		

Figura 38 – Cálculo de n₀ para o estrato 1 a 4 s.m. – 18 a 25 anos com E₀ = 2,5%

	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
4	ESTRATOS				N	%	e0	n0	n	n arred	
5	1 a 4 s.m.	18 a 25 anos	15	374	360	7,23%	0,025	1600	=(AB5*AE5)/(AB5+AE5)		
6	1 a 4 s.m.	25 a 40 anos	375	2053	1679	33,74%	0,025	1600			
7	1 a 4 s.m.	Mais de 40 anos	2054	2406	353	7,09%	0,025	1600			
8	4 a 12 s.m.	18 a 25 anos	2412	2486	75	1,51%	0,025	1600			
9	4 a 12 s.m.	25 a 40 anos	2487	3572	1086	21,82%	0,025	1600			
10	4 a 12 s.m.	Mais de 40 anos	3573	4205	633	12,72%	0,025	1600			
11	Mais de 12 s.m.	25 a 40 anos	4211	4551	341	6,85%	0,025	1600			
12	Mais de 12 s.m.	Mais de 40 anos	4552	5000	449	9,02%	0,025	1600			
13					4976	100,00%					

Figura 39 – Cálculo de n para o estrato 1 a 4 s.m. – 18 a 25 anos com E₀ = 2,5%

Mas, os valores de n são fracionários (Figura 42), precisam ser inteiros. Em cálculo de tamanho de amostra *sempre* arredondamos para cima. Podemos usar a função do Calc ARREDONDAR.PARA.CIMA na célula AG5, através do assistente de funções, resultando na Figura 40.

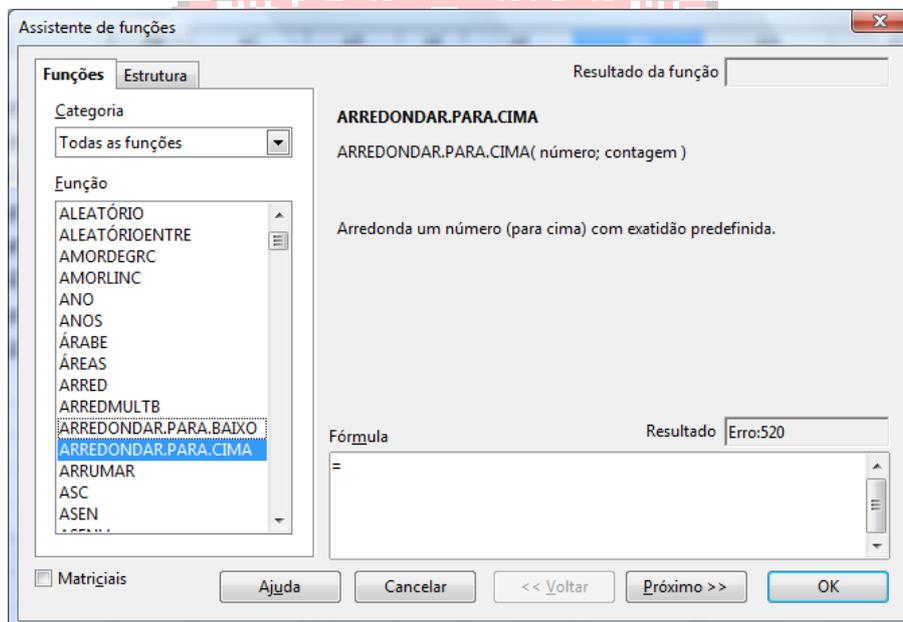


Figura 40 – Função ARREDONDAR.PARA.CIMA

Pressionando “Próximo” na Figura 40 chegamos às opções da Figura 41: em “número” devemos pôr a célula com o valor fracionário de n (célula AF5, para o primeiro estrato) e em “contagem” o número deve ser 0 (zero), indicando que o arredondamento será para número inteiro.

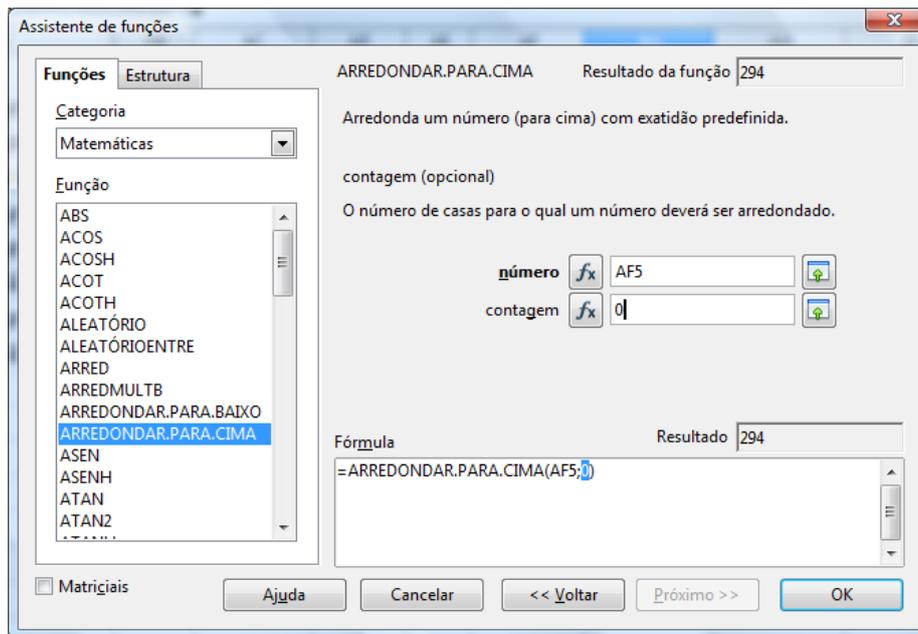


Figura 41 – Função ARREDONDAR.PARA.CIMA - Opções

Arrastando a célula AG5 até a AG12 obtemos os tamanhos de amostra para todos os estratos, e somando os valores chegamos ao tamanho total (Figura 42).

	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
4	ESTRATOS				N	%	e0	n0	n	n arred
5	1 a 4 s.m.	18 a 25 anos	15	374	360	7,23%	0,025	1600	293,88	294
6	1 a 4 s.m.	25 a 40 anos	375	2053	1679	33,74%	0,025	1600	819,27	820
7	1 a 4 s.m.	Mais de 40 anos	2054	2406	353	7,09%	0,025	1600	289,2	290
8	4 a 12 s.m.	18 a 25 anos	2412	2486	75	1,51%	0,025	1600	71,64	72
9	4 a 12 s.m.	25 a 40 anos	2487	3572	1086	21,82%	0,025	1600	646,91	647
10	4 a 12 s.m.	Mais de 40 anos	3573	4205	633	12,72%	0,025	1600	453,56	454
11	Mais de 12 s.m.	25 a 40 anos	4211	4551	341	6,85%	0,025	1600	281,09	282
12	Mais de 12 s.m.	Mais de 40 anos	4552	5000	449	9,02%	0,025	1600	350,61	351
13					4976	100,00%				3210

Figura 42 – Tamanhos de amostra para os estratos com $E_0 = 2,5\%$

Observe o tamanho total da amostra: 3210 elementos. Como a população tem 5000 elementos, poderia ser cogitado aplicar um censo diretamente... Os custos seriam altos. Torna-se atrativa a idéia de calcular um tamanho geral de amostra e dividi-la proporcionalmente entre os estratos.

2.2.2 – Tamanho de amostra com erro amostral definido para a população

Neste caso o n_0 é calculado uma única vez, para o erro amostral (E_0 definido). Corrige-se a estimativa com o tamanho total da população, e em seguida calculam-se os tamanhos de amostra de cada estrato de forma proporcional. Por exemplo, o estrato 1 a 4 s.m. – 18 a 25 anos corresponde a 7,23% da população, então ele corresponderá a 7,23% da amostra.

O procedimento citado acima tem um inconveniente: *o erro amostral dentro de cada estrato será maior do que o erro amostral “geral” usado na estimativa de n_0* . Se tivermos consciência disso ao calcular o tamanho de amostra e escrever um relatório a respeito, não haverá problema algum.

O cálculo de n_0 para E_0 igual a 2,5% (0,025) pode ser visto na Figura 43.

	X	Y	Z	AA	AB	AC	AD	AE
15	ESTRATOS		Início Fim	N	%		GERAL	
16	1 a 4 s.m.	18 a 25 anos	15 374	360	7,23%	e0		0,025
17	1 a 4 s.m.	25 a 40 anos	375 2053	1679	33,74%	n0		=1/AE16^2
18	1 a 4 s.m.	Mais de 40 anos	2054 2406	353	7,09%			
19	4 a 12 s.m.	18 a 25 anos	2412 2486	75	1,51%			
20	4 a 12 s.m.	25 a 40 anos	2487 3572	1086	21,82%			
21	4 a 12 s.m.	Mais de 40 anos	3573 4205	633	12,72%			
22	Mais de 12 s.m.	25 a 40 anos	4211 4551	341	6,85%			
23	Mais de 12 s.m.	Mais de 40 anos	4552 5000	449	9,02%			
24				4976	100,00%			

Figura 43 – Cálculo de n_0 para $E_0 = 2,5\%$ (toda a população)

Na Figura 44 calcula-se n usando o tamanho da população (4976).

	X	Y	Z	AA	AB	AC	AD	AE	AK	AL
15	ESTRATOS		Início Fim	N	%		GERAL			
16	1 a 4 s.m.	18 a 25 anos	15 374	360	7,23%	e0		0,025		
17	1 a 4 s.m.	25 a 40 anos	375 2053	1679	33,74%	n0		1600		
18	1 a 4 s.m.	Mais de 40 anos	2054 2406	353	7,09%	n		= (AB24*AE17)/(AB24+AE17)		
19	4 a 12 s.m.	18 a 25 anos	2412 2486	75	1,51%					
20	4 a 12 s.m.	25 a 40 anos	2487 3572	1086	21,82%					
21	4 a 12 s.m.	Mais de 40 anos	3573 4205	633	12,72%					
22	Mais de 12 s.m.	25 a 40 anos	4211 4551	341	6,85%					
23	Mais de 12 s.m.	Mais de 40 anos	4552 5000	449	9,02%					
24				4976	100,00%					

Figura 44 – Cálculo de n para $E_0 = 2,5\%$ (toda a população)

Novamente, precisamos arredondar o tamanho de amostra para cima (Figura 45), e posteriormente obter os tamanhos proporcionais, já arredondando para cima também (Figura 46).

	X	Y	Z	AA	AB	AC	AD	AE	AF	AJ	AK	AL	AM
15	ESTRATOS		Início Fim	N	%		GERAL						
16	1 a 4 s.m.	18 a 25 anos	15 374	360	7,23%	e0		0,025					
17	1 a 4 s.m.	25 a 40 anos	375 2053	1679	33,74%	n0		1600					
18	1 a 4 s.m.	Mais de 40 anos	2054 2406	353	7,09%	n		1210,7056					
19	4 a 12 s.m.	18 a 25 anos	2412 2486	76	1,53%	n arred		1211					
20	4 a 12 s.m.	25 a 40 anos	2487 3572	1085	21,80%								
21	4 a 12 s.m.	Mais de 40 anos	3573 4205	633	12,72%								
22	Mais de 12 s.m.	25 a 40 anos	4211 4551	341	6,85%								
23	Mais de 12 s.m.	Mais de 40 anos	4552 5000	449	9,02%								
24				4976	100,00%								

Figura 45 - Arredondamento para cima do valor de n

	X	Y	Z	AA	AB	AC	AD	AE	AF	AJ	AK	AL	AM
15	ESTRATOS		Início Fim	N	%		GERAL		Estrato				
16	1 a 4 s.m.	18 a 25 anos	15 374	360	7,23%	e0		0,025	=ARREDONDAR.PARA.CIMA(AC16*\$AE\$19;0)				
17	1 a 4 s.m.	25 a 40 anos	375 2053	1679	33,74%	n0		1600					
18	1 a 4 s.m.	Mais de 40 anos	2054 2406	353	7,09%	n		1210,7056					
19	4 a 12 s.m.	18 a 25 anos	2412 2486	75	1,51%	n arred		1211					
20	4 a 12 s.m.	25 a 40 anos	2487 3572	1086	21,82%								
21	4 a 12 s.m.	Mais de 40 anos	3573 4205	633	12,72%								
22	Mais de 12 s.m.	25 a 40 anos	4211 4551	341	6,85%								
23	Mais de 12 s.m.	Mais de 40 anos	4552 5000	449	9,02%								
24				4976	100,00%								

Figura 46 - Arredondamento para cima do valor de n no primeiro estrato

Arrastando a célula AF16 até a AF23 e posteriormente somando os tamanhos de amostra chegamos à Figura 47.

	X	Y	Z	AA	AB	AC	AD	AE	AF
15	ESTRATOS		Início	Fim	N	%	GERAL		Estrato
16	1 a 4 s.m.	18 a 25 anos	15	374	360	7,23%	e0	0,025	88
17	1 a 4 s.m.	25 a 40 anos	375	2053	1679	33,74%	n0	1600	409
18	1 a 4 s.m.	Mais de 40 anos	2054	2406	353	7,09%	n	1210,7056	86
19	4 a 12 s.m.	18 a 25 anos	2412	2486	75	1,51%	narred	1211	19
20	4 a 12 s.m.	25 a 40 anos	2487	3572	1086	21,82%			265
21	4 a 12 s.m.	Mais de 40 anos	3573	4205	633	12,72%			155
22	Mais de 12 s.m.	25 a 40 anos	4211	4551	341	6,85%			83
23	Mais de 12 s.m.	Mais de 40 anos	4552	5000	449	9,02%			110
24					4976	100,00%			1215

Figura 47 – Tamanhos de amostra para cada estrato

Observe que o total da amostra é de 1215 elementos, consideravelmente abaixo dos 3210 obtidos na seção 2.2.1. A redução de custos pode compensar o aumento do erro amostral dentro de cada estrato. Claro que se usarmos valores maiores de erro amostral o tamanho de amostra será menor, mas perderemos precisão. O pesquisador precisa balancear com cuidado custos de amostragem e precisão necessária: admite-se um erro amostral máximo de 5% (que acarreta um intervalo com 10% de extensão, o que pode ser demais em alguns casos).

Passaremos agora à obtenção da amostra propriamente dita.

2.2.3 – Obtenção da amostra estratificada proporcional

Vamos usar os tamanhos de amostra obtidos na seção 2.2.2 (1215 elementos no total). Na planilha Amostra.Estrat.Proporcional do arquivo PopulaçãoToyordSemErrosClassificado.ods é possível ver as colunas “Ordem” e “Sorteado” (Figura 48).

	A	B	C
1	Ordem	Sorteado	Modelo
2	1		#N/A
3	2		#N/A
4	3		#N/A
5	4		#N/A

Figura 48 – Colunas Ordem e Sorteado

Na coluna Ordem vemos em diferentes cores os componentes das amostras de cada estrato. Por exemplo, o estrato 1 a 4 s.m. – 18 a 25 anos compreende os sorteados de Ordem 1 a 88, cujas respectivas células estão em verde. Na célula B2 vamos pôr o primeiro elemento sorteado, através da função ALEATÓRIOENTRE (Figura 49).

	A	B	C	Opc
1	Ordem	Sorteado	Modelo	
2	1	=ALEATÓRIOENTRE(15;374)		
3	2		#N/A	
4	3		#N/A	
5	4		#N/A	

Figura 49 – Sorteio do primeiro elemento do primeiro estrato

Na Figura 47 podemos ver que o primeiro estrato está entre os casos 15 a 374 da planilha Dados do arquivo PopulaçãoToyordSemErrosClassificado.ods (que foi classificado na seção 1.3). Então, o sorteio precisa ser feito entre estes valores, como visto na Figura 49 ao lado.

Arrastando a célula B2 até a linha onde está ordem 88 chegamos à Figura 50.

	A	B	C	D
85	84	97	Chiconaultla	Inexistentes
86	85	322	Chiconaultla	Inexistentes
87	86	307	Chiconaultla	Inexistentes
88	87	303	Chiconaultla	Ar_e_direção
89	88	234	Valentiniana	Ar_e_direção

Figura 50 - Resultados da amostra do primeiro estrato - final⁵

A amostra do primeiro estrato foi selecionada. Agora podemos passar para o segundo estrato, por um procedimento análogo. O segundo estrato (ver Figura 47) vai do caso 375 ao 2053, o que resulta na Figura 51.

	A	B	C	D
89	88	211	Deltaforce3	AD_Trio_Elétrico
90	89	=ALEATÓRIOENTRE(375;2053)		
91	90		#N/A	#N/A

Figura 51 – Sorteio do primeiro elemento do segundo estrato

Arrastando a célula B90 até a linha onde está a ordem 497 ($497 = 88 + 409$, onde 409 é o tamanho da amostra calculada para o segundo estrato – Figura 47), células em azul celeste, completa-se a amostra do segundo estrato.

Devemos prosseguir este procedimento, tomando o cuidado de observar a delimitação dos casos de cada estrato em cada população (para preencher os argumentos das funções ALEATÓRIOENTRE) e a ordem dos elementos de cada estrato (para saber até onde arrastar as células com a função ALEATÓRIOENTRE). Veja o resumo dos casos e ordens dos estratos na Figura 52.

	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	
15			Casos população								Ordem	
16	ESTRATOS		Início	Fim	N	%	GERAL		Estrato	Inicial	Final	
17	1 a 4 s.m.	18 a 25 anos	15	374	360	7,23%	e0	0,025	88	1	88	
18	1 a 4 s.m.	25 a 40 anos	375	2053	1679	33,74%	n0	1600	409	89	497	
19	1 a 4 s.m.	Mais de 40 anos	2054	2406	353	7,09%	n	1210,7056	86	498	583	
20	4 a 12 s.m.	18 a 25 anos	2412	2486	75	1,51%	nared	1211	19	584	602	
21	4 a 12 s.m.	25 a 40 anos	2487	3572	1086	21,82%			265	603	867	
22	4 a 12 s.m.	Mais de 40 anos	3573	4205	633	12,72%			155	868	1022	
23	Mais de 12 s.m.	25 a 40 anos	4211	4551	341	6,85%			83	1023	1105	
24	Mais de 12 s.m.	Mais de 40 anos	4552	5000	449	9,02%			110	1106	1215	
25					4976	100,00%			1215			

Figura 52 – Resumo dos estratos (casos na população) e ordem de sorteio (na amostra)

Na planilha Amostra.Estrat.Proporcional do arquivo PopulaçãoToyordSemErros Classificado.ods é possível ver que as células correspondentes às ordens das amostras de cada estrato estão em cores diferentes para facilitar sua identificação.

⁵ Como o procedimento é aleatório o leitor pode obter resultados diferentes ao retirar sua amostra.