

Procedimentos de Análise Exploratória de Dados utilizando o Br.Office Calc ®
Professor Marcelo Menezes Reis

ÍNDICE ANALÍTICO

1) Procedimentos para variáveis qualitativas.....	3
1.1 Tabela de frequências e gráfico da variável Modelo (uma variável qualitativa).....	3
1.2 Tabela de frequências e gráfico do cruzamento das variáveis Modelo e Opinião Geral sobre o veículo (duas variáveis qualitativas).....	14
2. Procedimentos para Variáveis Quantitativas.....	19
2.1 - Procedimentos para variáveis quantitativas discretas.....	20
2.2 - Procedimentos para variáveis quantitativas contínuas.....	24
2.2.1 - Tabela de frequências agrupada em classes.....	24

ÍNDICE DE FIGURAS

Figura 1 – Assistente de dados.....	4
Figura 2 - Assistente de dados: origem dos dados.....	4
Figura 3 – Escolha das variáveis.....	4
Figura 4 – Layout da tabela dinâmica para a variável Modelo – 1ª parte.....	5
Figura 5 - Ações possíveis na tabela dinâmica.....	5
Figura 6 – Layout da tabela dinâmica para a variável Modelo – 2ª parte.....	6
Figura 7 – Layout da tabela dinâmica para a variável Modelo – 3ª parte.....	6
Figura 8 – Layout da tabela dinâmica para a variável Modelo – 4ª parte.....	7
Figura 9 – Tabela dinâmica para a variável Modelo.....	7
Figura 10 – Cálculo de percentuais para a variável Modelo.....	8
Figura 11 – Tabela dinâmica para a variável Modelo (contagem e percentuais).....	8
Figura 12 – Ícone de gráfico do Calc.....	8
Figura 13 – Assistente gráfico do Calc – 1ª parte.....	9
Figura 14 – Assistente gráfico do Calc – 2ª parte.....	9
Figura 15 – Assistente gráfico do Calc – 2ª parte – modificação da seleção.....	10
Figura 16 – Assistente gráfico do Calc – 3ª parte.....	10
Figura 17 – Assistente gráfico do Calc – 4ª parte.....	10
Figura 18 – Gráfico para a variável Modelo.....	11
Figura 19 – Opções do gráfico para a variável Modelo.....	11
Figura 20 – Edição do gráfico para a variável Modelo – 1ª parte.....	11
Figura 21 – Edição do gráfico para a variável Modelo – 2ª parte.....	12
Figura 22 – Gráfico em pizza para a variável Modelo – 1ª parte.....	12
Figura 23 – Gráfico em pizza para a variável Modelo – 2ª parte.....	12
Figura 24 – Gráfico em pizza para a variável Modelo – 2ª parte.....	13
Figura 25 – Gráfico em pizza para a variável Modelo – 3ª parte.....	13
Figura 26 - Layout da tabela dinâmica de Modelo x Opinião Geral.....	14
Figura 27 – Tabela de contingências Modelo x Opinião Geral.....	14
Figura 28 – Modificação da tabela de contingências Modelo x Opinião Geral – 1ª parte.....	15
Figura 29 – Modificação da tabela de contingências Modelo x Opinião Geral – 2ª parte.....	15
Figura 30 – Modificação da tabela de contingências Modelo x Opinião Geral – 3ª parte.....	15
Figura 31 – Modificação da tabela de contingências Modelo x Opinião Geral – 4ª parte.....	16
Figura 32 – Tabela de contingências Modelo x Opinião Geral – % de linha.....	16
Figura 33 – Tabela de contingências Modelo x Opinião Geral – % de linha – mudança de formato.....	16
Figura 34 – Tabela de contingências Modelo x Opinião Geral – % de linha.....	17
Figura 35 – Assistente gráfico para Modelo x Opinião Geral – 2ª parte.....	17
Figura 36 – Assistente gráfico para Modelo x Opinião Geral – 3ª parte.....	17
Figura 37 – Assistente gráfico para Modelo x Opinião Geral – 4ª parte.....	18
Figura 38 – Modelo x Opinião Geral – gráfico de barras múltiplas - frequências.....	18
Figura 39 – Mudança do gráfico de barras múltiplas – frequências para percentagem empilhada.....	19
Figura 40 – Modelo x Opinião Geral – gráfico de barras múltiplas – percentagem empilhada por modelo.....	19
Figura 41 – Tabela de frequências -Anos de remodelação.....	20
Figura 42 - Tabela de Frequência: Anos de educação.....	20
Figura 43 – Assistente gráfico do Calc – 1ª parte.....	21
Figura 44 – Assistente gráfico do Calc – 2ª parte.....	21

Figura 45 – Assistente gráfico do Calc – 3ª parte.....	21
Figura 46 – Assistente gráfico do Calc – 4ª parte.....	22
Figura 47 – Gráfico de anos de remodelação.....	22
Figura 48 – Opções do gráfico para a variável Anos de remodelação.....	22
Figura 49 – Edição do gráfico para a variável Anos de remodelação – 1ª parte.....	23
Figura 50 – Edição do gráfico para a variável Anos de remodelação – 2ª parte.....	23
Figura 51 – Edição do gráfico para a variável Anos de remodelação – 3ª parte.....	23
Figura 52 – Histograma de Anos de remodelação.....	24
Figura 53 – Funções disponíveis no Calc ®.....	24
Figura 54 – Funções disponíveis no Calc ® por categoria.....	25
Figura 55 – Funções estatísticas disponíveis no Calc ®.....	25
Figura 56 - Função Mínimo.....	26
Figura 57 - Valores mínimo e máximo de Renda.....	26
Figura 58 - Amplitude de classes de Renda.....	27
Figura 59 – Cálculo dos limites inferior e do limite superior da primeira classe de Renda.....	27
Figura 60 – Cálculo dos limites inferior e superior da segunda classe de Renda.....	27
Figura 61 - Classes de Renda.....	28
Figura 62 - Ponto médio da primeira classe de Renda.....	28
Figura 63 - Obtenção da Frequência da primeira classe de Renda.....	29
Figura 64 - Obtenção da Frequência da última classe de Renda.....	29
Figura 65 - Obtenção da Frequência acumulada da primeira classe de Renda.....	30
Figura 66 - Obtenção da Frequência acumulada da primeira classe de Renda.....	30
Figura 67 - Renda agrupada em classes: frequências absolutas, percentuais, frequências acumuladas e percentuais acumulados.....	31
Figura 68 - Histograma dos salários anuais (agrupados em classes).....	32
Figura 67 - Função lógica SE.....	33
Figura 68 - Argumentos da função SE.....	34
Figura 69 - Procedimento para categorizar a variável Renda.....	35
Figura 70 - Variável Renda Categorizada.....	36
Figura 71 - Tabela dinâmica da variável Renda Categorizada.....	36
Figura 72 - Medidas de síntese da variável Renda.....	37
Figura 73 – Menu Dados: opção Classificar.....	38
Figura 74 – Critérios de classificação.....	38
Figura 75 – Critérios de classificação – escolha de variáveis.....	39
Figura 76 – Base de dados ordenada em função de Modelo.....	39
Figura 77 – Medidas de síntese de Renda em função de Modelo; medidas gerais de Renda.....	40
Figura 78 – Assistente gráfico do Calc – diagrama de dispersão – 1ª parte.....	40
Figura 79 – Assistente gráfico do Calc – diagrama de dispersão – 2ª parte.....	41
Figura 80 – Assistente gráfico do Calc – diagrama de dispersão – 3ª parte.....	41
Figura 81 – Assistente gráfico do Calc – diagrama de dispersão – 3ª parte - seleção.....	41
Figura 82 – Assistente gráfico do Calc – diagrama de dispersão – 3ª parte – dados prontos.....	42
Figura 83 – Assistente gráfico do Calc – diagrama de dispersão – 4ª parte.....	42
Figura 84 – Quilometragem por Renda.....	42
Figura 85 – Opções do gráfico para o diagrama de dispersão Quilometragem por Renda.....	43
Figura 86 - Seleção do eixo Y.....	43
Figura 87 - Opções de formatação de eixo: escala.....	43
Figura 88 - Formatação de eixo: escala modificada.....	44
Figura 89 - Diagrama de dispersão Quilometragem por Renda.....	44

O objetivo deste texto é apresentar os principais procedimentos de Análise Exploratória de Dados, tal como apresentados nas Unidades 3 e 4, mas utilizando a planilha eletrônica Calc (da suíte Br.Office. Os dados estão no arquivo AmostraToyord, disponível no ambiente virtual: contém as informações sobre 250 consumidores de uma montadora fictícia, a Toyord, selecionados por amostragem probabilística. Há 11 variáveis: as 10 variáveis realmente pesquisadas (que estão descritas no Exemplo 1 da Unidade 3) e o número do caso sorteado da população original, que tinha 5000 elementos. Todos as figuras elaboradas pelo autor, adaptando de Sun Microsystems.

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. Estes são
- 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.

1) Procedimentos para variáveis qualitativas.

Quando se deseja realizar uma análise exploratória em que apenas variáveis qualitativas estejam envolvidas, o mais usual é construir tabelas de frequências para cada variável individualmente, ou tabelas de contingências (dupla classificação) relacionando duas variáveis. Gráficos podem ser produzidos a partir das tabelas. Para lidar com variáveis qualitativas no Calc precisamos utilizar as tabelas dinâmicas. Veja os exemplos a seguir:

1.1 Tabela de frequências e gráfico da variável Modelo (uma variável qualitativa).

a) Procure pelo menu "Dados" na barra principal do Calc. Procure pela opção "Assistente de dados", e o comando "Iniciar", veja a figura 1. Escolha esta opção.

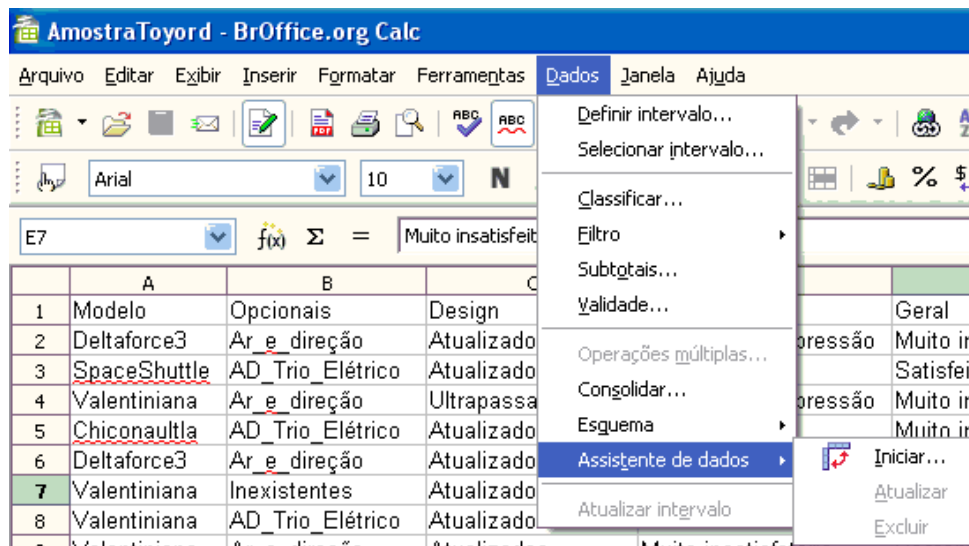


Figura 1 – Assistente de dados

b) Ao escolher esta opção surgirá uma tela parecida com a figura 2:

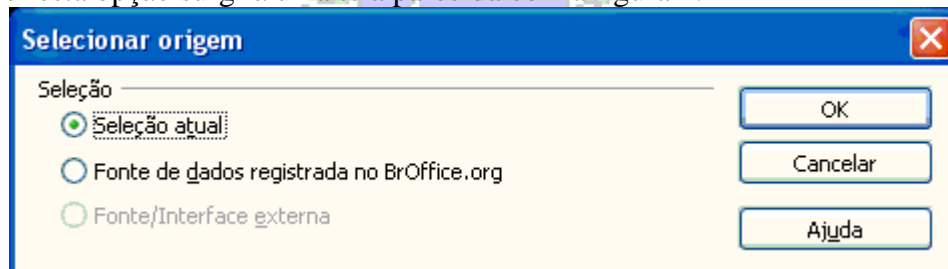


Figura 2 - Assistente de dados: origem dos dados

c) Automaticamente o Calc pode selecionar o conjunto de variáveis na planilha aberta, no caso as 10 variáveis de interesse. Caso não seja este o conjunto de dados que você quer analisar, pode escolher outro intervalo de dados através da opção “Fonte de dados registrada no BrOffice.org”. Mas, aqui pretendemos trabalhar com as variáveis selecionadas, pressionando OK, chegamos na figura 3, onde vemos o layout da tabela e as variáveis disponíveis.

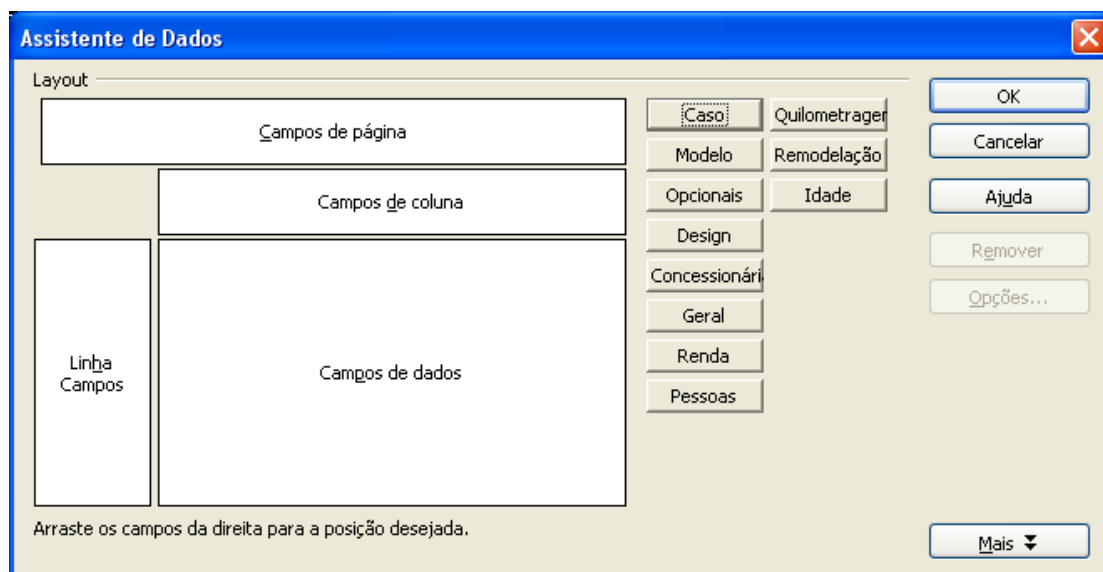


Figura 3 – Escolha das variáveis

d) Observe à direita os nomes das variáveis existentes no arquivo de dados, e que foram selecionadas na etapa anterior. Como há interesse apenas na variável Modelo devemos selecioná-la e arrastá-la até o campo "LINHA", ou "COLUNA". As outras variáveis não farão parte da tabela. Devemos também arrastar a variável Modelo para o campo "DADOS". Isso é necessário para especificar a ação que a tabela deve executar. A ação padrão é a soma dos valores, tal como mostrado na figura 4.

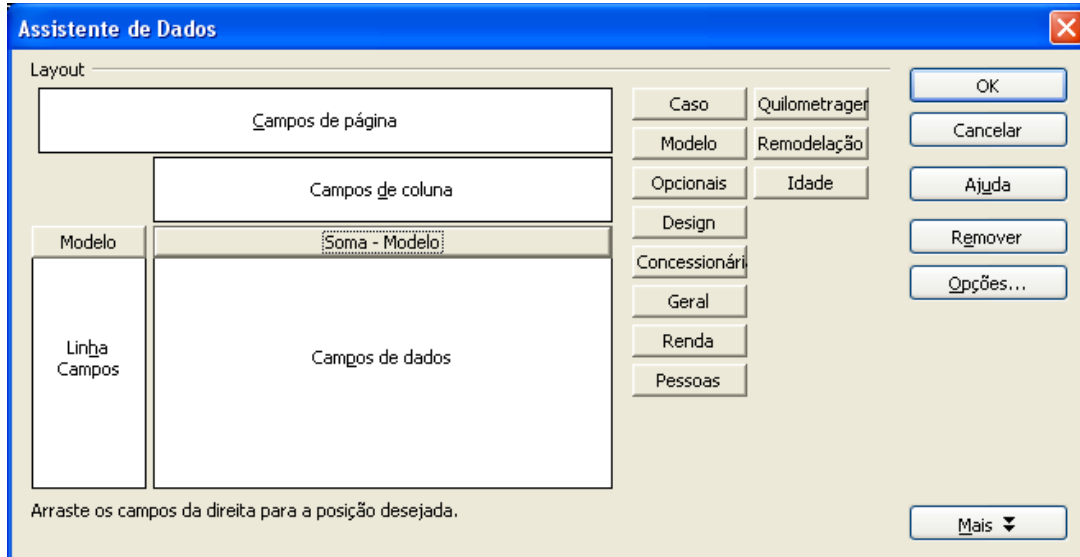


Figura 4 – Layout da tabela dinâmica para a variável Modelo – 1ª parte

Nós não queremos a soma de modelo. Mesmo porque não faz sentido, pois modelo é qualitativa... Precisamos da contagem do número de ocorrências de cada categoria de modelo para analisar a variável. Se clicarmos duas vezes sobre “Soma – Modelo”, vamos obter a figura 5:

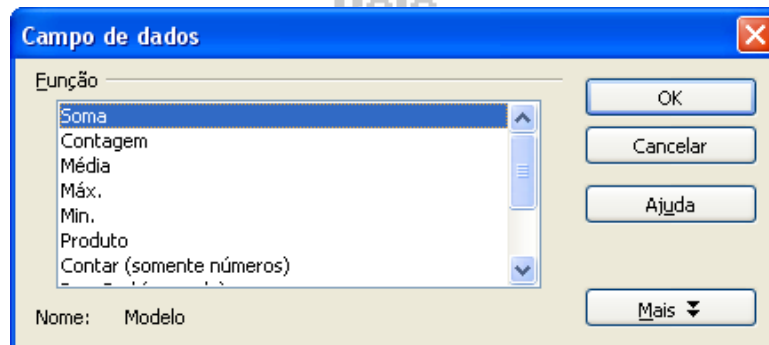


Figura 5 - Ações possíveis na tabela dinâmica

Na figura acima estão as várias ações possíveis (como média, mínimo, máximo, desvio padrão, que serão úteis quando formos analisar variáveis quantitativas), e a opção “Contagem”. No presente caso queremos contar o número de ocorrências dos valores da variável Modelo, portanto selecionamos “Contagem”, e pressionamos "OK" (na tela da figura 5). O resultado é mostrado na figura 6:

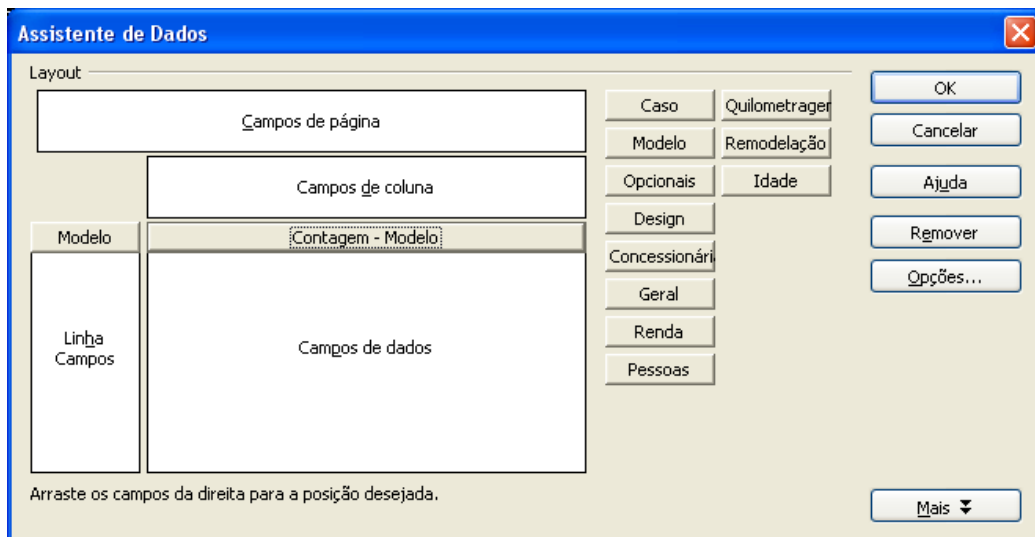


Figura 6 – Layout da tabela dinâmica para a variável Modelo – 2ª parte

Precisamos agora definir onde será posto o resultado da tabela. Pressionando “Mais” na tela da Figura 6 chegaremos ao resultado da figura 7.

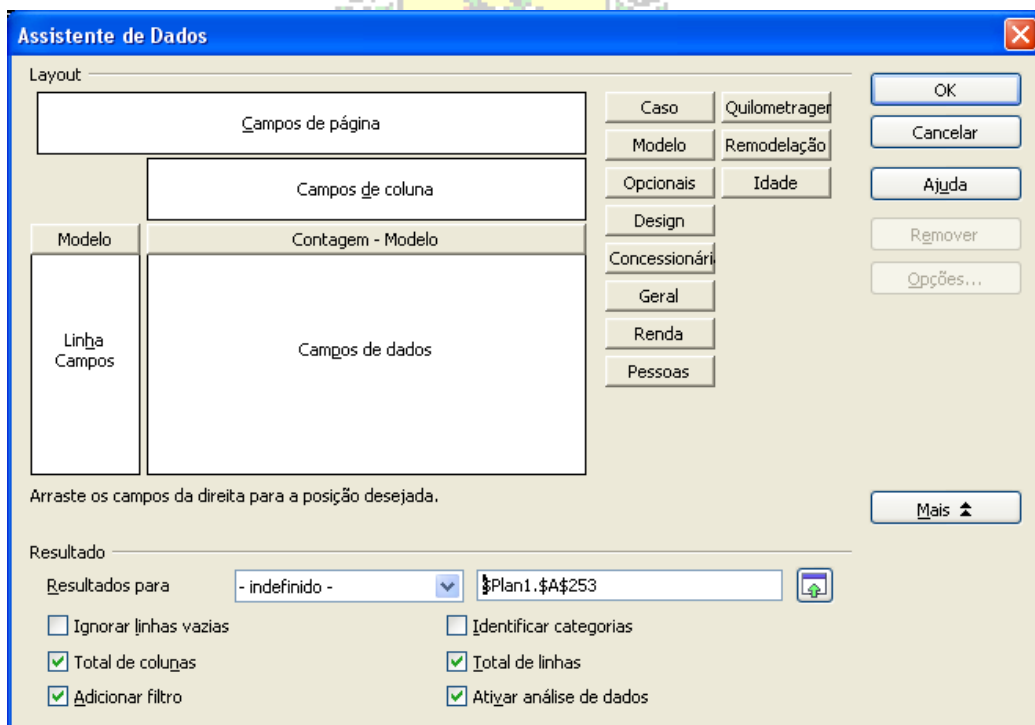


Figura 7 – Layout da tabela dinâmica para a variável Modelo – 3ª parte

e) O padrão do Calc é pôr os resultados a partir de uma célula pouco abaixo da última que contém dados, neste caso o Calc selecionou a célula A253. Observe que as opções “Total de colunas” e “Total de linhas” foram selecionadas também, bem como “Adicionar filtro” e “Ativar análise de dados”. Mas, vamos pôr os resultados em uma nova planilha, para não congestionar a planilha de dados. Assim, no campo “Resultados para” vamos selecionar “Nova planilha”, levando à figura 8.

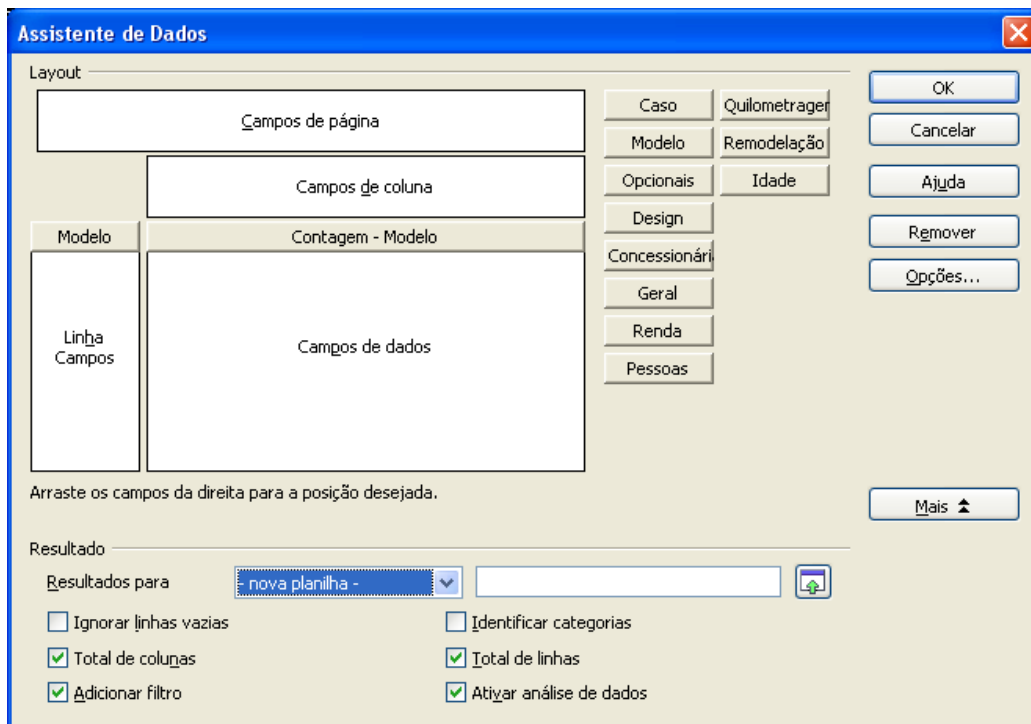


Figura 8 – Layout da tabela dinâmica para a variável Modelo – 4ª parte

Pressionando OK chegamos ao resultado da tabela dinâmica:

	A	B	C
1	Filtro		
2			
3	Modelo		
4		1	
5	Chiconaultla	81	
6	Deltaforce3	56	
7	LuxuriousCar	29	
8	SpaceShuttle	42	
9	Valentiniana	41	
10	Total Result	250	

Figura 9 – Tabela dinâmica para a variável Modelo

f) Observe que a tabela contém as frequências para os cinco modelos de veículos, além do campo vazio, pois um dos elementos da amostra não respondeu qual era o seu modelo de veículo. Infelizmente o Calc não permite pôr na mesma tabela as contagens absolutas e os percentuais (o Excel permite...), o que nos força a obtê-los manualmente, através de fórmulas na planilha. Basta dividir cada Frequência pelo total geral (célula B10), e multiplicar o resultado por 100. Veja a figura 10.

	A	B	C
1	Filtro		
2			
3	Modelo		Percentual
4		1	$= (B4/\$B\$10)*100$
5	Chiconaultla	81	
6	Deltaforce3	56	
7	LuxuriousCar	29	
8	SpaceShuttle	42	
9	Valentiniana	41	
10	Total Resultado	250	

Figura 10 – Cálculo de percentuais para a variável Modelo

É muito importante fazer com que o denominador da expressão (o total geral, que está na célula B10) seja uma referência absoluta, para que ao “arrastar” a fórmula para as outras categorias de Modelo a divisão seja sempre feita por 250 (que é o total geral). Fazemos isso pondo o símbolo \$ antes da letra (que define a coluna) e do número (que define a linha): \$B\$10. Tomando este cuidado e “arrastando” a fórmula até a célula C10 vamos obter:

	A	B	C
1	Filtro		
2			
3	Modelo		Percentual
4		1	0,4
5	Chiconaultla	81	32,4
6	Deltaforce3	56	22,4
7	LuxuriousCar	29	11,6
8	SpaceShuttle	42	16,8
9	Valentiniana	41	16,4
10	Total Resultado	250	100

Figura 11 – Tabela dinâmica para a variável Modelo (contagem e percentuais)

g) Com a tabela pronta podemos avaliar os resultados. É possível também construir gráficos com base nela. Para tanto devemos pressionar o ícone “Gráfico” na barra de ferramentas superior do Calc:

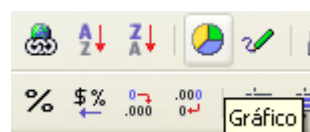


Figura 12 – Ícone de gráfico do Calc

Ao pressionar “Gráfico” surgirá a caixa com as opções da figura 13:

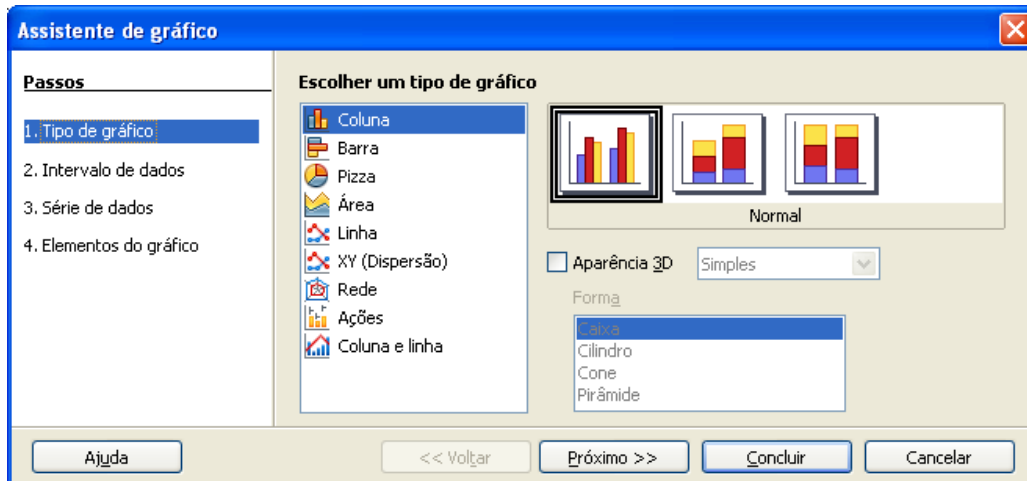


Figura 13 – Assistente gráfico do Calc – 1ª parte

Primeiramente precisamos escolher o tipo de gráfico. Como se trata de uma variável qualitativa temos aqui três opções: coluna (barras verticais, o gráfico de barras visto na unidade 3), barra (barras horizontais), e pizza (em setores, visto também na unidade 3). Vamos escolher “Coluna”, mantendo a primeira opção (padrão). Pressionando “Próximo” chegamos à tela da figura 14:

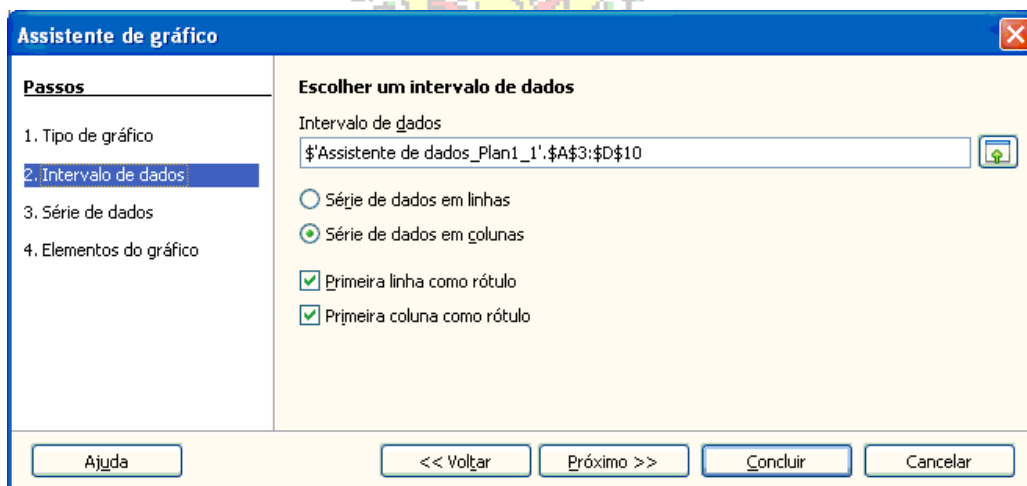


Figura 14 – Assistente gráfico do Calc – 2ª parte

Devemos escolher o intervalo de dados. Basta selecionar a tabela dinâmica e o Calc automaticamente irá registrar o intervalo A3:D10 (incluindo a coluna com os percentuais e mais uma coluna vazia). Contudo, não devemos incluir as colunas C e D no gráfico, e, tampouco a linha 10, pois esta contém o total geral, que vai apenas “poluir” o gráfico. Então selecionamos apenas o intervalo A3:B9, como na figura 15:

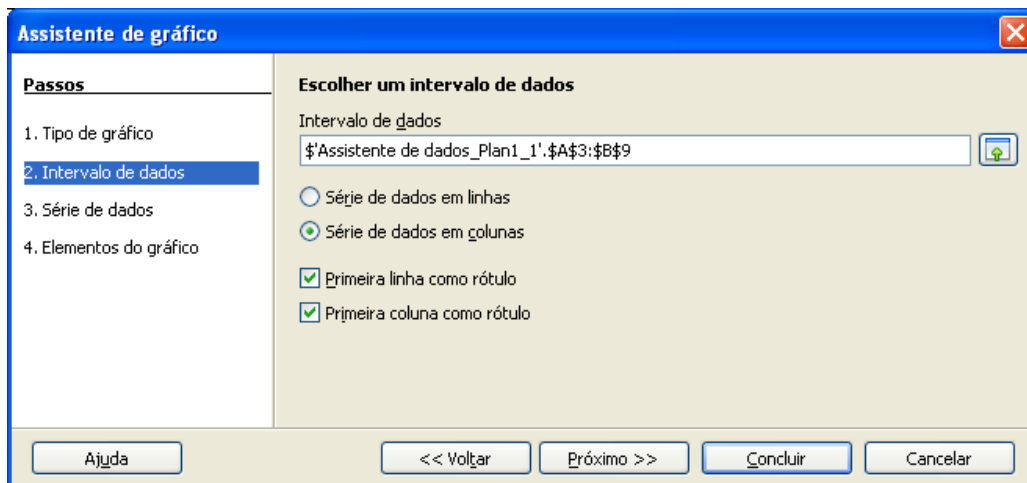


Figura 15 – Assistente gráfico do Calc – 2ª parte – modificação da seleção

Podemos pressionar “Próximo” e chegar à figura 16:

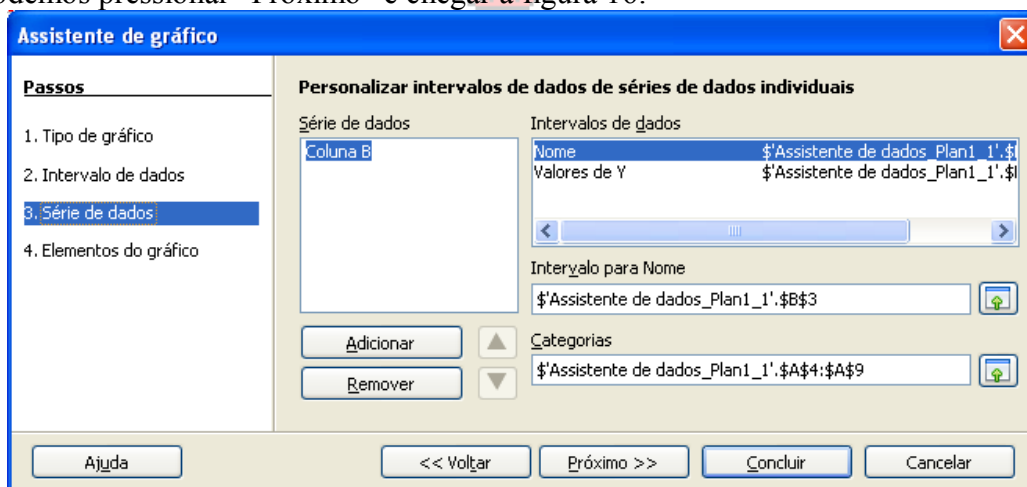


Figura 16 – Assistente gráfico do Calc – 3ª parte

Não é preciso fazer mudanças aqui. A série de dados está correta e as categorias também. Basta pressionar “Próximo” para chegar à última etapa do gráfico.

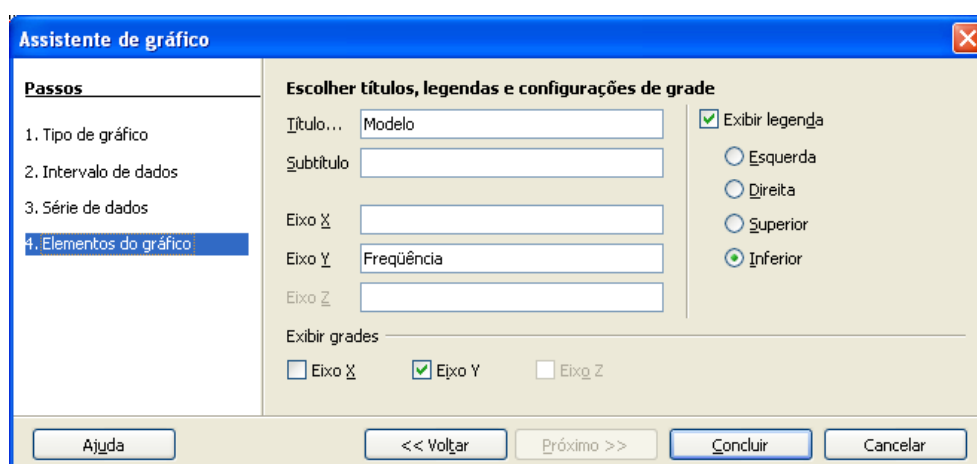


Figura 17 – Assistente gráfico do Calc – 4ª parte

Podemos escolher o título do gráfico, se haverá exibição de legenda, os nomes dos eixos, etc. Vamos chamar o gráfico de “Modelo”, e no eixo vertical estará “Frequência”, e a legenda será exibida no canto inferior. Podemos, então, pressionar “Concluir”, e o resultado será:

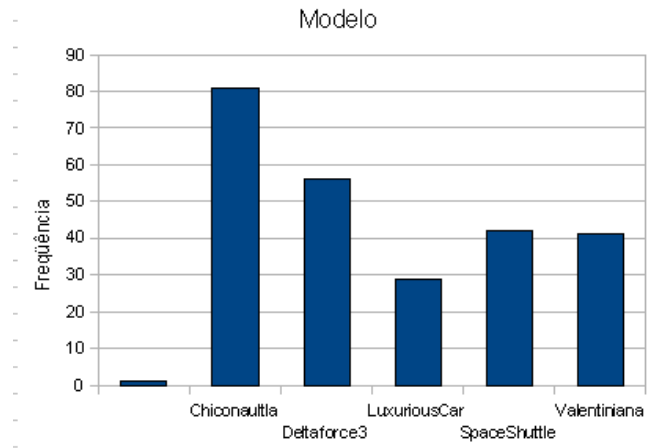


Figura 18 – Gráfico para a variável Modelo

Com o gráfico pronto podemos fazer a interpretação: o modelo Chiconautla é o mais vendido, seguido por DeltaForce3, SpaceShuttle, Valentiniana e LuxuriousCar.

Do ponto de vista das frequências o gráfico está pronto, mas podemos modificar aspectos estéticos, como título, escala, cor de fundo, cores das barras, entre outros aspectos. Para modificar o gráfico selecione-o e pressione o botão direito do mouse, levando à figura 19:

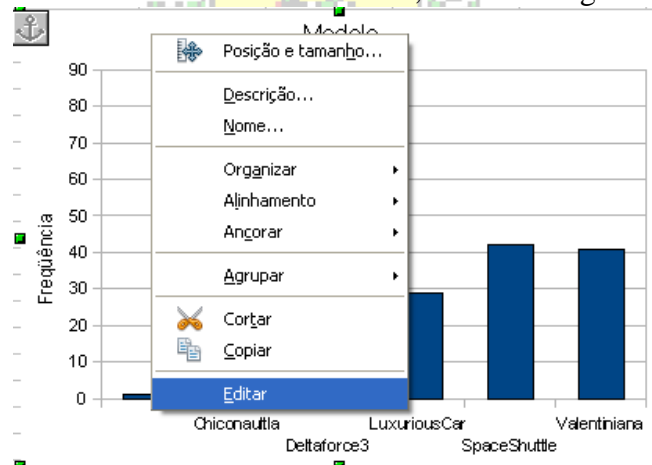


Figura 19 – Opções do gráfico para a variável Modelo

Dentre as opções há “Editar”. Através dela podemos modificar mesmo o intervalo dos dados do gráfico, e o tipo de gráfico. Ao pressionar “Editar” o gráfico fica emoldurado por uma linha cinza, e ao pôr o cursor sobre ele, e clicando o botão direito do mouse, chegamos à figura 20:

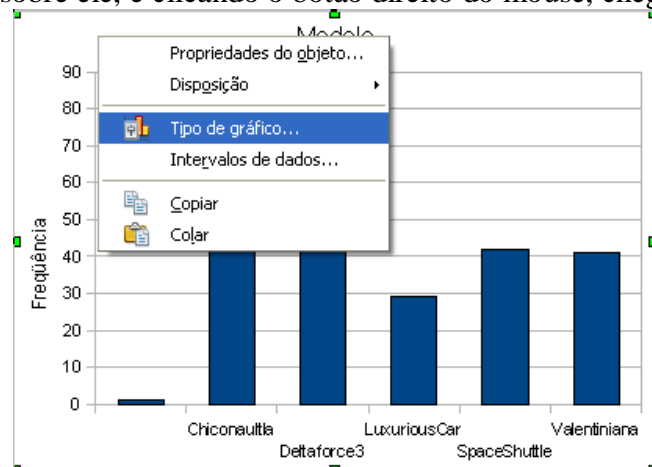


Figura 20 – Edição do gráfico para a variável Modelo – 1ª parte

Vamos mudar o tipo de gráfico (para pizza, também chamado de gráfico em setores, por exemplo). Selecionando “Tipo de gráfico” voltamos à tela da figura 13, reproduzida na figura 21:

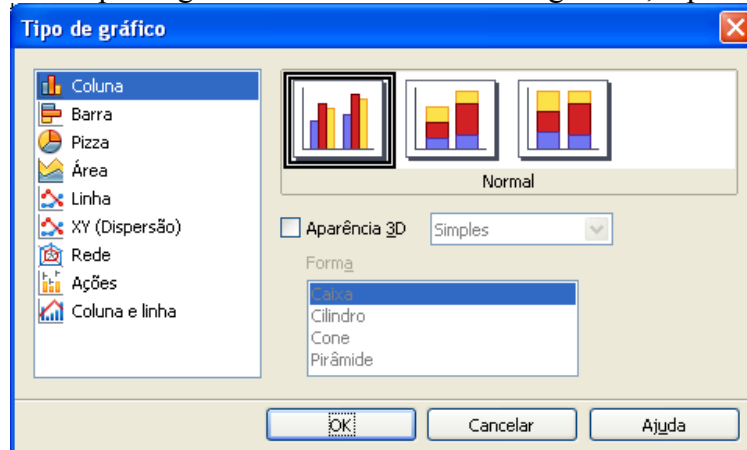


Figura 21 – Edição do gráfico para a variável Modelo – 2ª parte

Selecionando “Pizza”, e mantendo a opção padrão, obtemos o gráfico mostrado na figura 22:

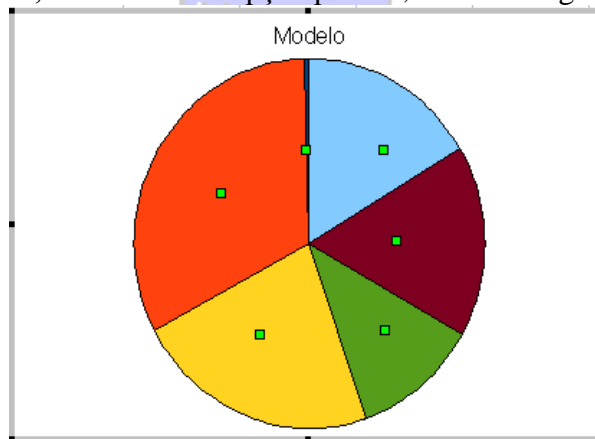


Figura 22 – Gráfico em pizza para a variável Modelo – 1ª parte

O gráfico da figura 22 não nos diz muita coisa, pois não há legenda ou percentuais. É preciso modificar as propriedades. Para tanto, selecione o gráfico com o mouse. Em seguida, pressione o botão direito do mouse, obtendo uma tela parecida com a da figura 19. Escolha a opção “Editar”: o gráfico ficará emoldurado por uma linha cinza. Ao pôr o cursor sobre o gráfico e pressionando o botão direito do mouse surgirá a tela da figura 23:

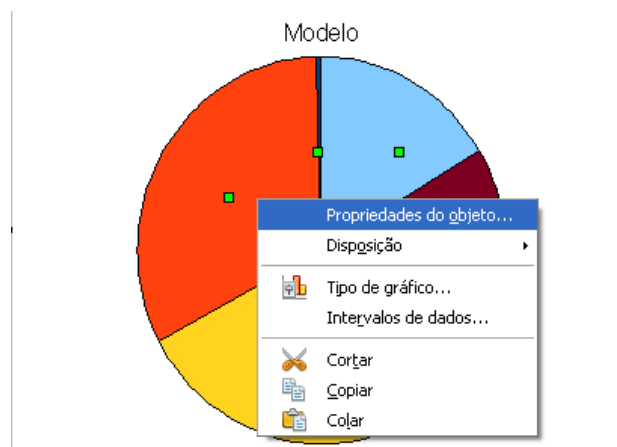


Figura 23 – Gráfico em pizza para a variável Modelo – 2ª parte

Selecionado “Propriedades do objeto” chegamos à figura 24:

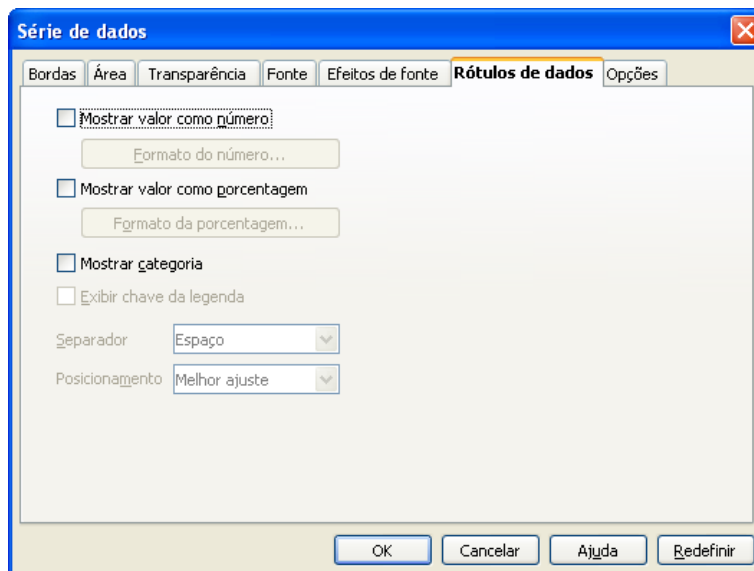


Figura 24 – Gráfico em pizza para a variável Modelo – 2ª parte

Devemos selecionar a opção “Rótulo de dados”. Para que a legenda seja mostrada é preciso marcar a opção “Mostrar categoria”, e o Calc permite que possamos mostrar os valores de contagem como número (“Frequência”) ou porcentagem. No gráfico em pizza faz mais sentido usar porcentagens. Assim, o resultado final será:

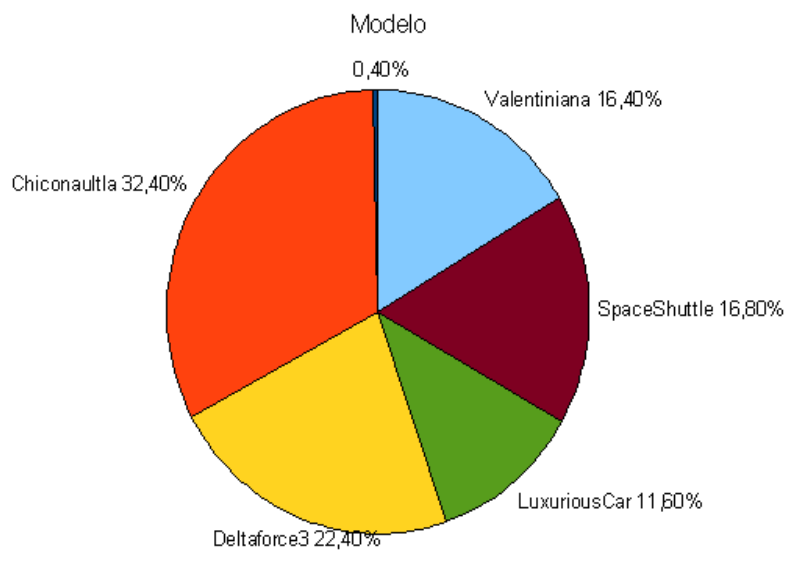


Figura 25 – Gráfico em pizza para a variável Modelo – 3ª parte

O tamanho das fontes da legenda, o formato das porcentagens, entre outros aspectos, podem ser modificados para melhorar a visualização.

A interpretação é muito semelhante a do gráfico em barras, mas agora utilizando percentuais diretamente: Chiconaultla é o modelo mais vendido com 33% do mercado, seguido por DeltaForce3 (22%). Valentiniiana e SpaceShuttle têm fatias de mercado bem semelhantes, em torno de 16%, e em último lugar LuxuriousCar com 12% do mercado.

1.2 Tabela de frequências e gráfico do cruzamento das variáveis Modelo e Opinião Geral sobre o veículo (duas variáveis qualitativas).

O procedimento é semelhante ao utilizado no item 1.1, mas agora serão utilizadas duas variáveis, havendo o interesse de construir uma tabela de contingências.

a) O procedimento descrito nas letras a) a c) do item 1.1 (figuras 1 a 3) pode ser repetido literalmente. As diferenças começam a aparecer na determinação do layout da tabela. Agora precisamos cruzar as variáveis Modelo e Opinião Geral (Geral): Modelo ficará no campo “Linha” e Geral no campo “Coluna”. No campo “Dados” poremos qualquer uma das duas, mas APENAS uma delas. O resultado, após modificar a opção de mostrar os dados como Soma para Contagem, está na figura 26 :

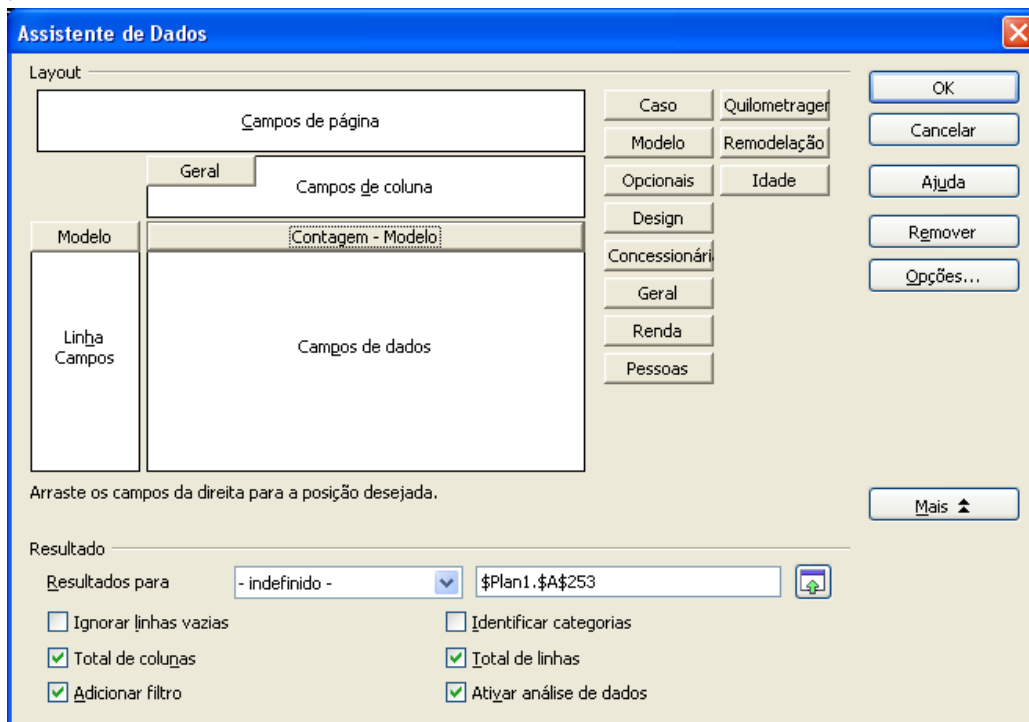


Figura 26 - Layout da tabela dinâmica de Modelo x Opinião Geral

Infelizmente o Calc não permite acrescentar percentuais à tabela de contingências. É preciso calculá-los à parte: percentuais em relação aos totais das linhas, colunas ou total geral. Conforme visto na Unidade 3 estes percentuais são imprescindíveis para avaliarmos o relacionamento entre as variáveis, e podem ser calculados usando a metodologia vista na letra f da seção 1.1.

Escolhemos pôr os resultados em uma nova planilha (ver letra d da seção 1.1), com os totais de linhas e colunas, e o resultado será:

	A	B	C	D	E	F
1	Filtro					
2						
3	Cont.Núm - Modelo	Geral				
4	Modelo	Bastante satisfeito	Insatisfeito	Muito insatisfeito	Satisfeito	Total Resultado
5				1		1
6	Chiconaultla		11	69	1	81
7	Deltaforce3		22	29	5	56
8	LuxuriousCar	19	1		9	29
9	SpaceShuttle	10	14	1	17	42
10	Valentiniana	3	18	11	9	41
11	Total Resultado	32	67	110	41	250

Figura 27 – Tabela de contingências Modelo x Opinião Geral

Observe que a primeira linha da tabela contém uma observação vazia para Modelo, mas que a opinião geral do cliente sobre este veículo não identificado é “Insatisfeito”. Para avaliar se há associação entre as variáveis a obtenção de percentuais é muito importante.

b) Podemos calcular os percentuais à parte, ou construir novamente a tabela pedindo para apresentar os resultados como percentuais (das linhas, colunas ou total geral).

Basta selecionar qualquer parte da tabela e pressionar o botão direito do mouse, resultando na figura 28:

	A	B	C	D	E	F
1	Filtro					
2						
3	Cont.Núm -	Geral				
4	Modelo	Bastante sat	Insatisfeito	Muito insatis	Satisfeito	Total Resultado
5						1
6	Chiconaultla			69	1	81
7	Deltaforce3			29	5	56
8	LuxuriousCar				9	29
9	SpaceShuttle			1	17	42
10	Valentiniana			11	9	41
11	Total Result			110	41	250
12						
13						
14						

Figura 28 – Modificação da tabela de contingências Modelo x Opinião Geral – 1ª parte

Selecionando “Iniciar” o Calc volta à tela mostrada na figura 26. Selecionando “Contagem-Modelo” o botão “Opções” torna-se visível. Pressionando-o, surge a tela da figura 29:

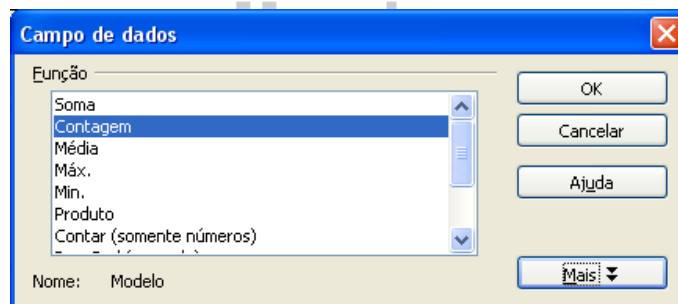


Figura 29 – Modificação da tabela de contingências Modelo x Opinião Geral – 2ª parte

Já vimos a figura 28 quando mudamos a opção de Soma para Contagem (figura 5). Agora, porém, vamos pressionar o botão “Mais”, que permitirá apresentar os resultados de uma forma diferente.

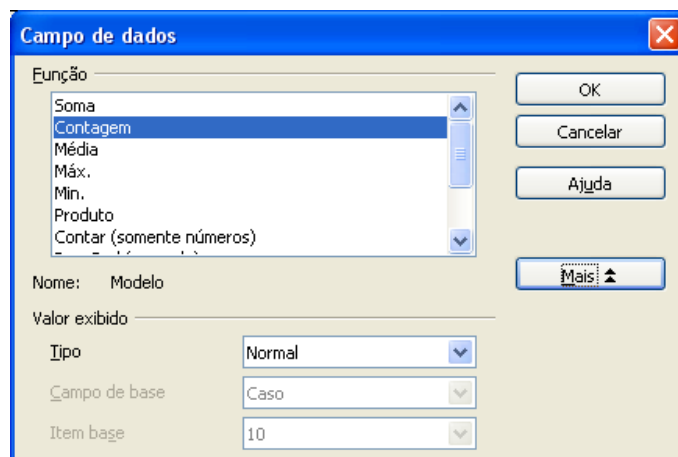


Figura 30 – Modificação da tabela de contingências Modelo x Opinião Geral – 3ª parte

O padrão é a exibição do valor como “Normal”, no nosso caso a Frequência absoluta da contagem. Mas, podemos mudar para % de linha, ou outras formas, como mostrado na figura 31:

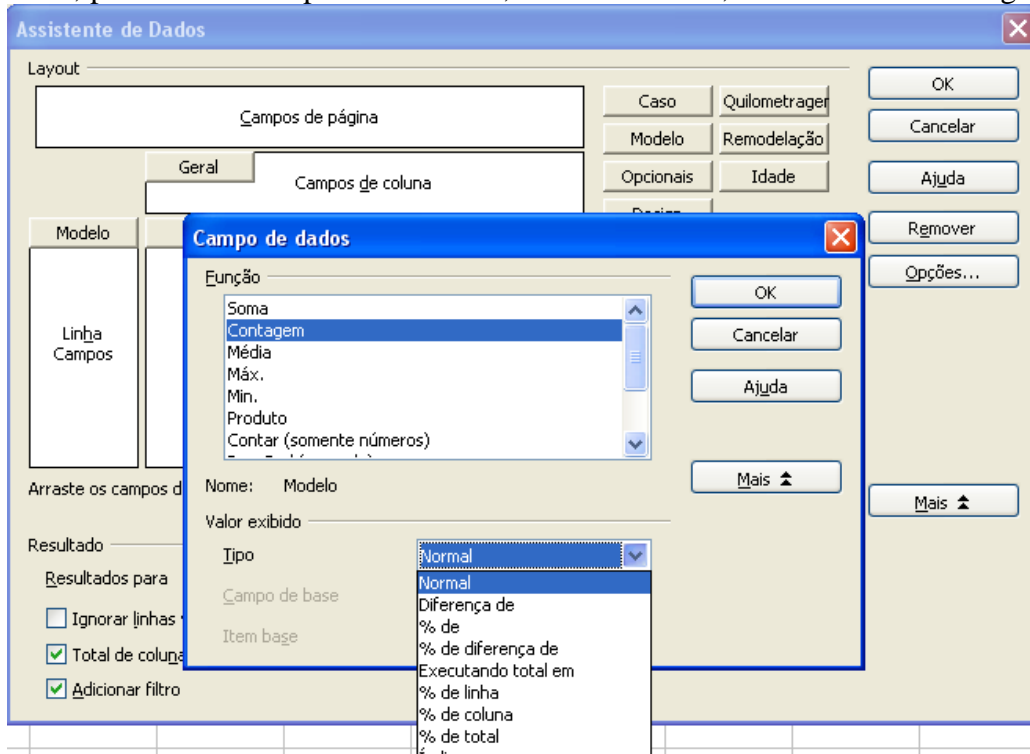


Figura 31 – Modificação da tabela de contingências Modelo x Opinião Geral – 4ª parte

Selecionando % de linha, e pressionando OK até fechar todos os diálogos, eis a figura 32:

	A	B	C	D	E	F
1	Filtro					
2						
3	Cont.Núm - Geral					
4	Modelo	Bastante satis	Insatisfeito	Muito insatis	Satisfeito	Total Resultado
5				1		1
6	Chiconaultla		0,14	0,85	0,01	1
7	Deltaforce3		0,39	0,52	0,09	1
8	LuxuriousCar	0,66	0,03		0,31	1
9	SpaceShuttle	0,24	0,33	0,02	0,4	1
10	Valentiniana	0,07	0,44	0,27	0,22	1
11	Total Resultado	0,13	0,27	0,44	0,16	1

Figura 32 – Tabela de contingências Modelo x Opinião Geral – % de linha

Observe que os dados estão em proporções absolutas, a soma das linhas resulta 1 (100%). Para mudar a apresentação para % é preciso selecionar todas as células da tabela e mudar o formato do número para % conforme mostrado na figura 33.

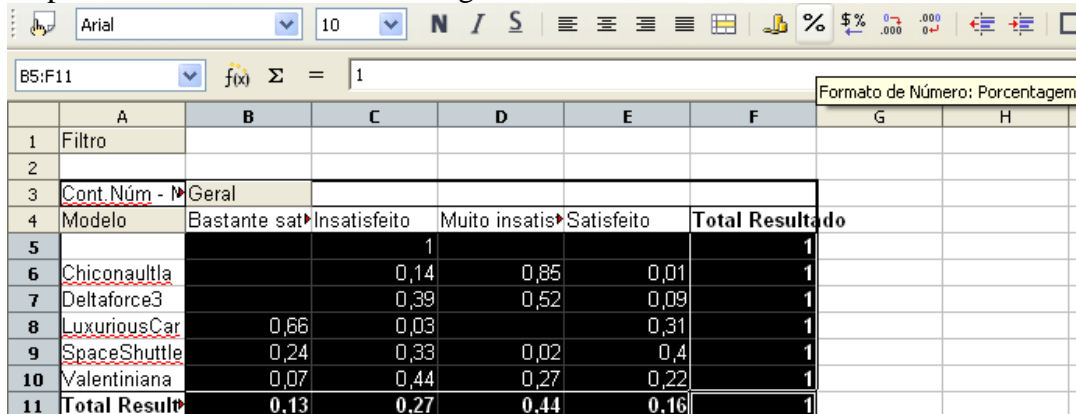


Figura 33 – Tabela de contingências Modelo x Opinião Geral – % de linha – mudança de formato

O resultado final, com as porcentagens por linha é mostrado na figura 34.

	A	B	C	D	E	F
1	Filtro					
2						
3	Cont.Núm - Modelo	Geral				
4	Modelo	Bastante satisfeito	Insatisfeito	Muito insatisfeito	Satisfeito	Total Resultado
5			100,00%			100,00%
6	Chiconaultla		13,58%	85,19%	1,23%	100,00%
7	Deltaforce3		39,29%	51,79%	8,93%	100,00%
8	LuxuriousCar	65,52%	3,45%		31,03%	100,00%
9	SpaceShuttle	23,81%	33,33%	2,38%	40,48%	100,00%
10	Valentiniana	7,32%	43,90%	26,83%	21,95%	100,00%
11	Total Resultado	12,80%	26,80%	44,00%	16,40%	100,00%

Figura 34 – Tabela de contingências Modelo x Opinião Geral – % de linha

Interpretando a tabela é possível ver a insatisfação com o Modelo Chiconaultla (85,19% dos seus proprietários declaram-se muito insatisfeitos) e a satisfação com o LuxuriousCar (65,52% declaram-se bastante satisfeitos). Há evidência de relacionamento entre as variáveis, pois dependendo do modelo a satisfação dos clientes é muito diferente.

c) Da mesma forma que para uma variável qualitativa podemos construir um gráfico de barras, mas aqui múltiplas, para representar a tabela da figura 27. Usando procedimentos análogos aos das figuras 12 a 18 (item g da seção 1.1) vamos ter:

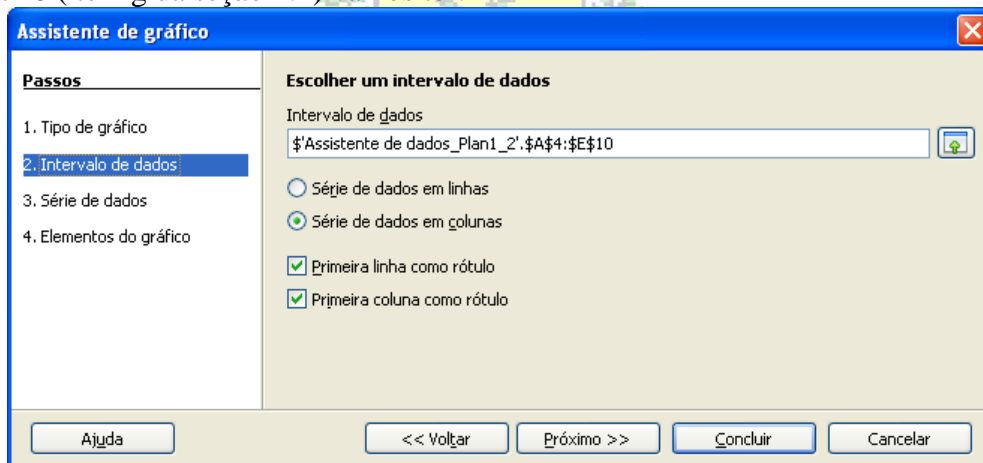


Figura 35 – Assistente gráfico para Modelo x Opinião Geral – 2ª parte

Tomamos o cuidado de selecionar apenas os dados da célula A4 a E10, deixando os totais de fora, pois eles irão apenas “poluir” o gráfico. Pressionando “Próximo” chegamos à figura 36:

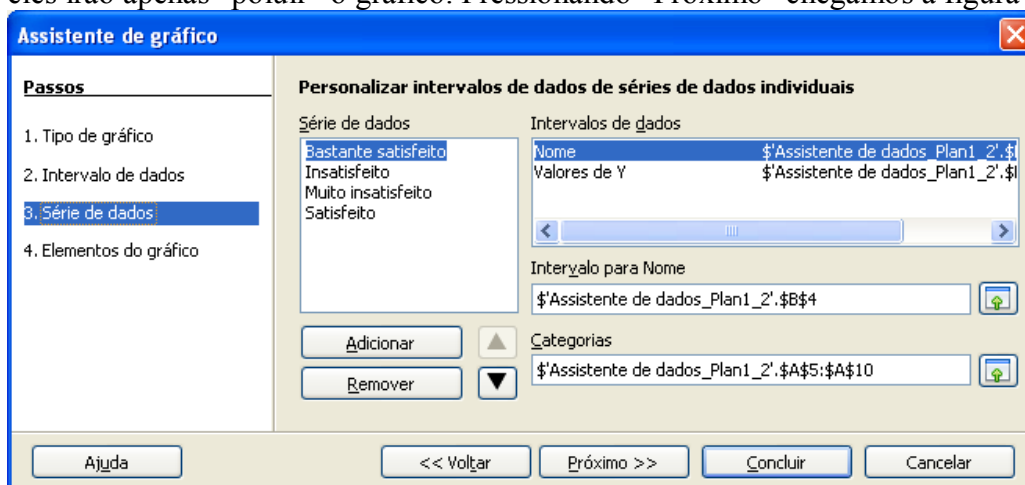


Figura 36 – Assistente gráfico para Modelo x Opinião Geral – 3ª parte

Pressionando “Próximo” novamente chegamos à figura 37:

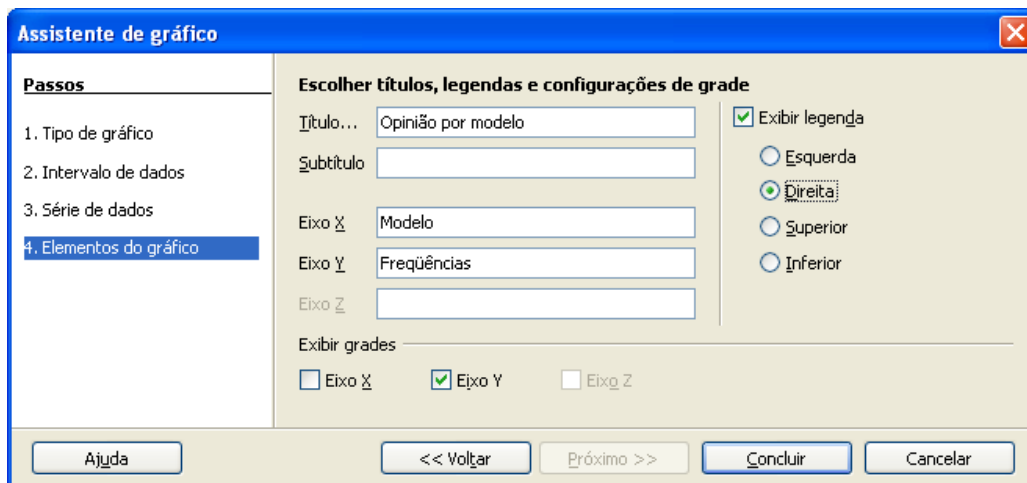


Figura 37 – Assistente gráfico para Modelo x Opinião Geral – 4ª parte

Aqui acrescentamos o título do gráfico e os nomes nos eixos, além de definir que a legenda ficará à direita. O resultado final está a seguir:

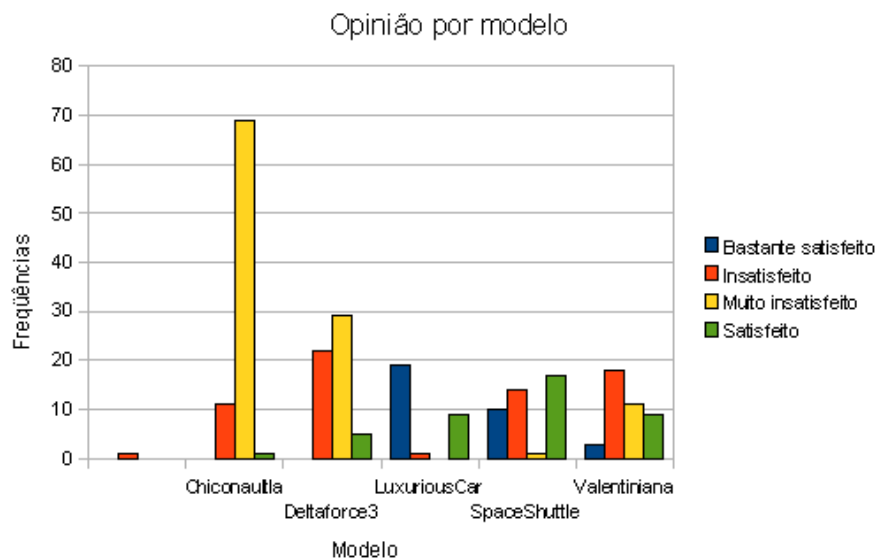


Figura 38 – Modelo x Opinião Geral – gráfico de barras múltiplas - frequências

Observe que o gráfico foi apresentado com as frequências dos cruzamentos. Mas a interpretação pode ser feita comparando as diferenças relativas entre as barras.

- As barras referentes a “Insatisfeito” e “Muito insatisfeito” são mais “altas” nos modelos Chiconaultla e DeltaForce3 (não há nenhuma observação “Bastante satisfeito” nos dois modelos...).
- À medida que passamos para os modelos mais sofisticados as frequências das opiniões negativas vão reduzindo e as das positivas vão aumentando (não há nenhuma observação “Muito insatisfeito” para o LuxuriousCar).

O gráfico da figura 38 poderia ser modificado para apresentar os percentuais (por modelo – por linha, como na figura 34). Basta mudar o tipo de gráfico, usando os procedimentos mostrados nas figuras 19 e 20. A tela na figura 39 mostra as opções:

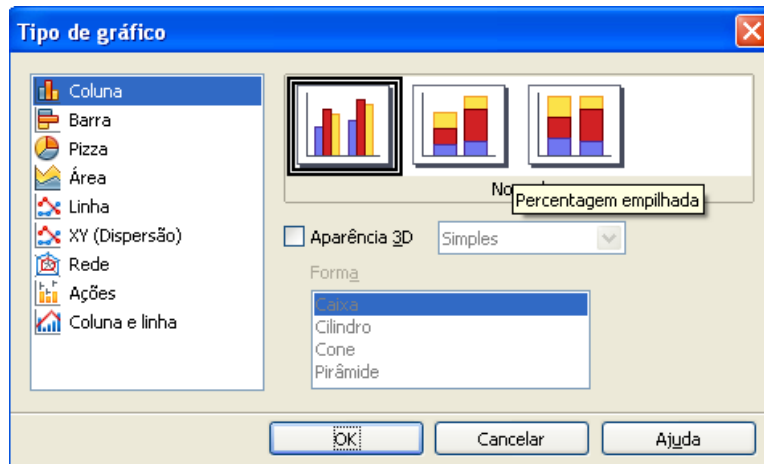


Figura 39 –Mudança do gráfico de barras múltiplas – frequências para percentagem empilhada

A opção “percentagem empilhada” levará ao resultado apresentado na figura 34, mas agora de forma gráfica, na figura 40:

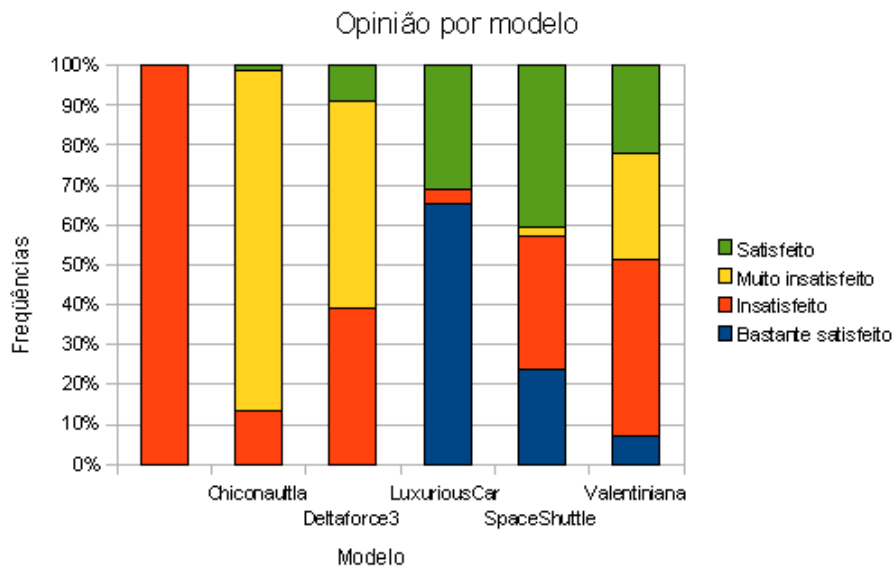


Figura 40 –Modelo x Opinião Geral – gráfico de barras múltiplas – percentagem empilhada por modelo

A mesma interpretação feita para a tabela da figura 34 pode ser feita aqui. Os modelos Chiconaultla e DeltaForce3 apresentam índices de insatisfação alarmantes, enquanto LuxuriousCar e SpaceShuttle foram aprovados por seus proprietários.

Vamos ver agora os procedimentos para variáveis quantitativas.

2. Procedimentos para Variáveis Quantitativas

Para variáveis quantitativas há uma maior variedade de procedimentos disponíveis. É preciso distinguir os procedimentos relativos às variáveis discretas e contínuas, bem como os casos em que queremos fazer uma análise de uma variável quantitativa em função de uma qualitativa (por exemplo, Renda por Modelo). Vamos ver os procedimentos nos próximos itens.

2.1 - Procedimentos para variáveis quantitativas discretas

Se a variável for discreta, por exemplo Remodelação (há quantos anos os clientes acham que os veículos foram remodelados) no arquivo “AmostraToyord.ods”, o procedimento pode ser semelhante ao utilizado para a variável Modelo (seção 1.1): ao construir a tabela dinâmica o Calc irá selecionar "Soma de Remodelação" como ação, e precisamos modificar isso para contagem dos valores: "Contar de Remodelação". Seguindo o procedimento do item 1.1, incluindo os percentuais para cada valor, mas agora para a variável Remodelação, vamos obter:

	A	B
1	Filtro	
2		
3	Remodelação	
4	0	2
5	1	57
6	2	123
7	3	59
8	4	9
9	Total Resultado	250

Figura 41 – Tabela de frequências -Anos de remodelação

É possível perceber que uma parcela significativa (123 em 250, mais de 75%) dos clientes acredita que os veículos foram remodelados há pelo menos 2 anos. Isso é preocupante para uma montadora de automóveis, que atua em um ramo extremamente competitivo, e que a aparência de modernidade de seus modelos é um importante diferencial. Mudanças são necessárias...

Da mesma forma que para Modelo (seção 1.1) podemos construir um gráfico. Mas, precisamos ter algum cuidado. Imagine que você obtivesse uma tabela como a da figura 45.

Contagem de AnosEd	Total
AnosEd	
8	53
12	190
14	6
15	116
16	59
17	11
18	9
19	27
20	2
21	1

Figura 42 - Tabela de Frequência: Anos de educação

NÃO construa um gráfico diretamente a partir desta tabela. Estão faltando valores da variável, de 9 a 11, e 13 (não há pessoas com esses anos de educação). Para construir um histograma para variáveis discretas, você tem que copiar os dados da tabela ao lado para outras células e acrescentar os valores 9 a 11, e 13, mantendo suas frequências iguais a zero. Voltemos ao caso da figura 44.

Como a variável anos de Remodelação não tem valores com Frequência nula, podemos construir o gráfico diretamente a partir da tabela dinâmica.

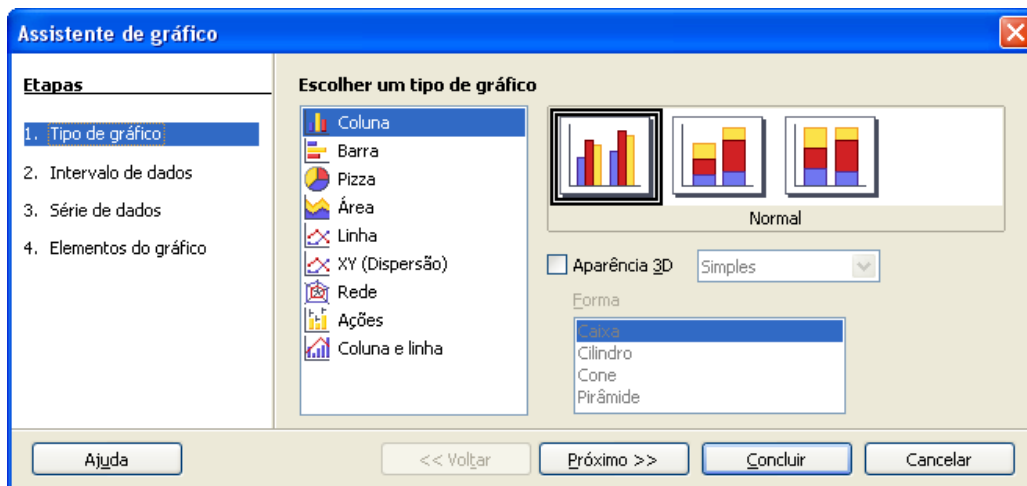


Figura 43 – Assistente gráfico do Calc – 1ª parte

No assistente gráfico não há a opção “Histograma”, mas podemos escolher o gráfico em colunas, pois um histograma não deixa de ser um gráfico de colunas justapostas. Ao fazer a escolha e pressionar “Próximo” chegaremos à opção “Intervalo de dados”. Se, como no nosso caso, a tabela dinâmica foi selecionada diretamente o próprio Calc ® preencherá o intervalo de dados. Mas, ao fazer isso ele inclui a linha 9, o que irá complicar o gráfico. Portanto, o intervalo de dados deverá ser da célula A3 a B8, como visto na figura 44:

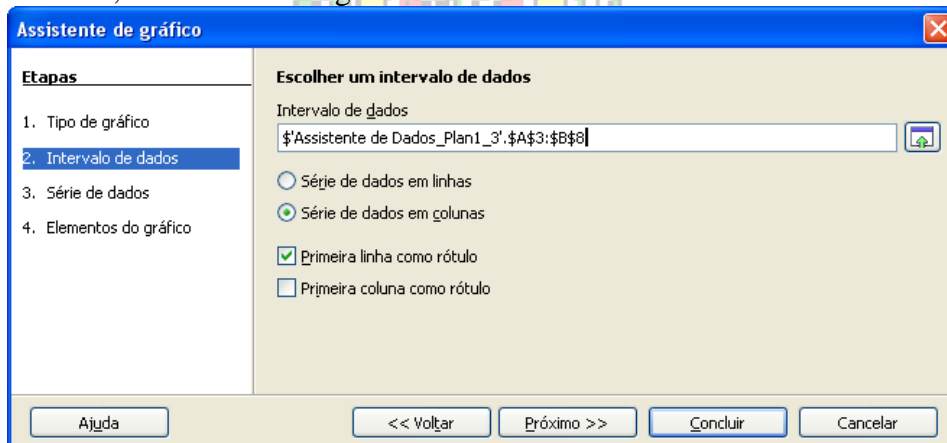


Figura 44 – Assistente gráfico do Calc – 2ª parte

Com os dados selecionados, basta pressionar “Próximo” e chegar à figura 45.

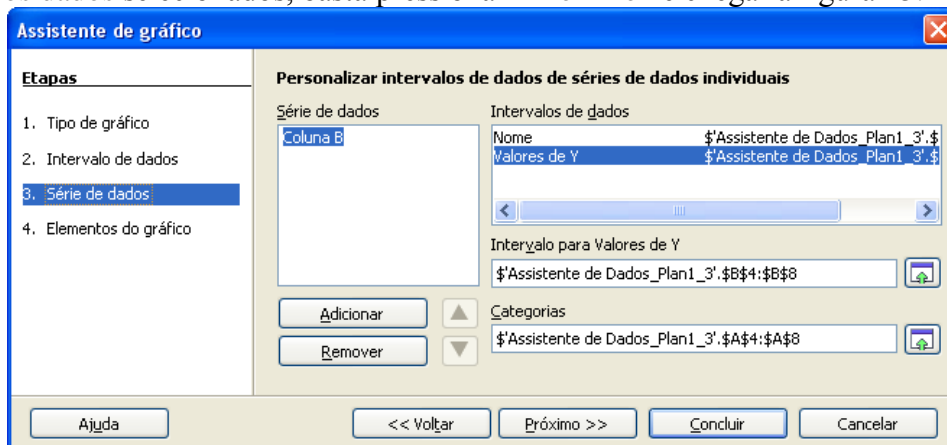


Figura 45 – Assistente gráfico do Calc – 3ª parte

Observe que o Calc ® chamou os dados de “Coluna B”, o que seria útil, se fosse um cruzamento entre variáveis. Mas não é este o caso, “Coluna B” será posta como legenda, o que não precisamos neste gráfico. Pressionando “Próximo” chegamos à última parte:

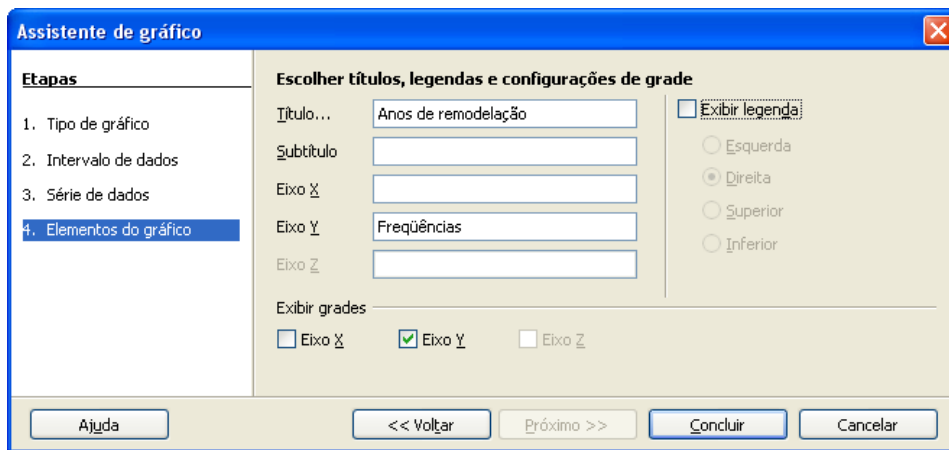


Figura 46 – Assistente gráfico do Calc – 4ª parte

Nesta etapa podemos modificar o título dos eixos, o título do próprio gráfico, a eventual colocação de legenda, entre outros aspectos. Vamos desmarcar a exibição da legenda. Posteriormente, basta pressionar “Concluir” e obter o gráfico da figura 47.

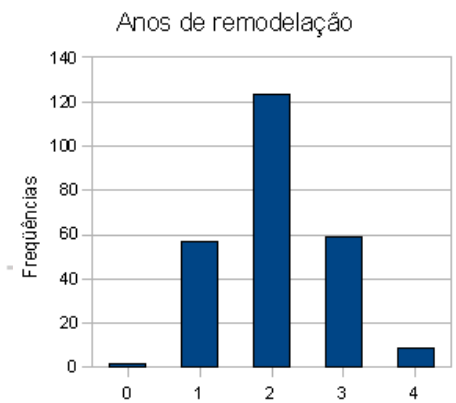


Figura 47 – Gráfico de anos de remodelação

Há algo estranho no histograma acima... As barras não são justapostas, pois o Calc ® tem como padrão inserir espaços entre as barras dos gráficos em colunas. Isso precisa ser corrigido. O gráfico da figura 47 NÃO É um histograma, pois as barras não estão justapostas. Para modificar o gráfico selecione-o e pressione o botão direito do mouse, levando à figura 48 (veja figura 18 também):

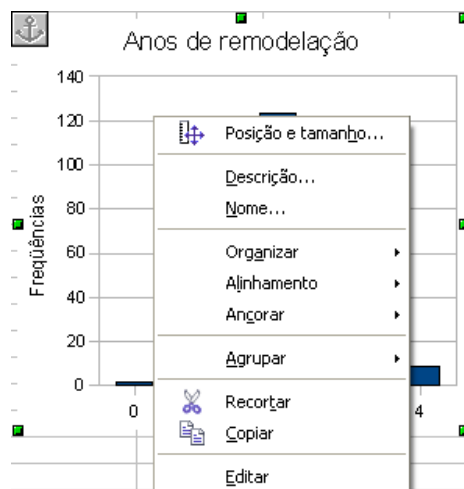


Figura 48 – Opções do gráfico para a variável Anos de remodelação

Dentre as opções disponíveis há “Editar”. Através dela podemos modificar mesmo o intervalo dos dados do gráfico, e o tipo de gráfico. Ao pressionar “Editar” o gráfico fica emoldurado por uma linha cinza, e ao pôr o cursor sobre ele, e clicando o botão direito do mouse, chegamos à figura 49:

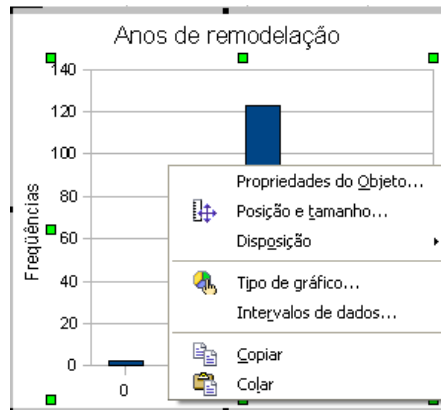


Figura 49 – Edição do gráfico para a variável Anos de remodelação – 1ª parte

Queremos mudar o espaçamento entre as barras. Selecionando “Propriedades do objeto” chegamos à figura 50:

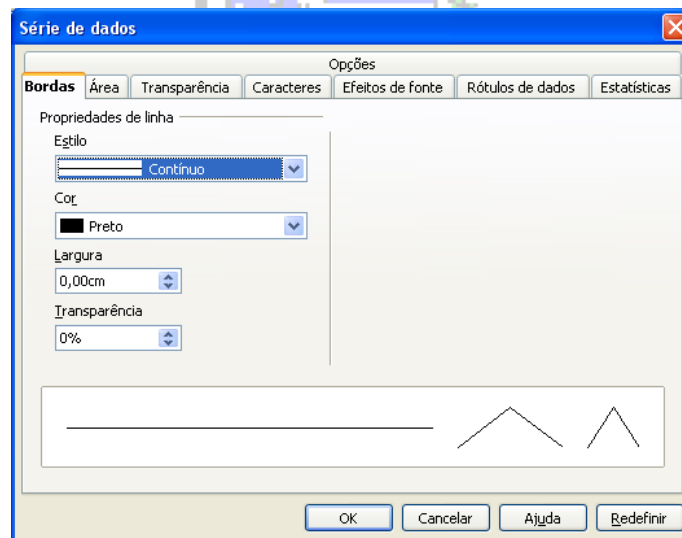


Figura 50 – Edição do gráfico para a variável Anos de remodelação – 2ª parte

É possível realizar todo tipo de alteração cosmética no gráfico. Mas, se escolhermos a aba “Opções” vamos encontrar o que procuramos.

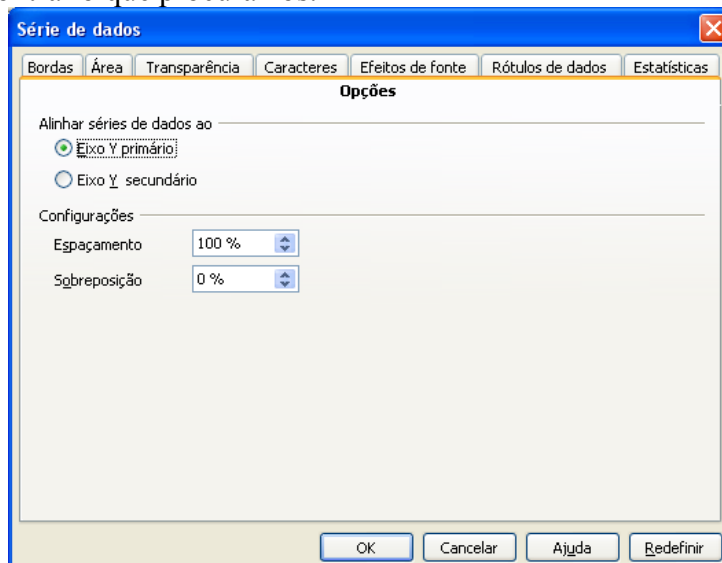


Figura 51 – Edição do gráfico para a variável Anos de remodelação – 3ª parte

Em “Configurações” há a opção “Espaçamento”, que está definida como 100% (padrão). Basta torná-la igual a zero, e o gráfico passa a ser um histograma correto (figura 52):

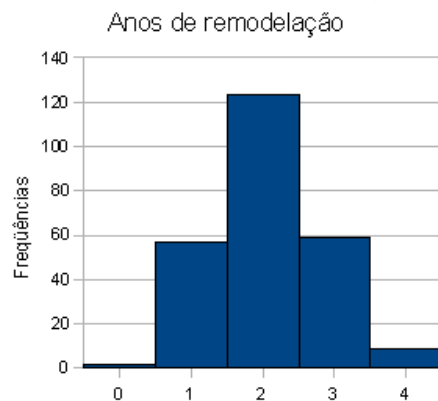


Figura 52 – Histograma de Anos de remodelação

Rapidamente percebemos que o valor 2 ocorre com maior Frequência.

2.2 - Procedimentos para variáveis quantitativas contínuas.

Se a variável de interesse for quantitativa contínua, como Renda mensal, "Renda" no arquivo "AmostraToyord.ods", pode haver interesse em construir uma tabela de frequências agrupada em classes, tal como demonstrado na Unidade 3. Para isso, NÃO PODEMOS utilizar a tabela dinâmica do Calc ®: como os valores de uma variável contínua repetem-se pouco (ou não se repetem), teríamos uma tabela imensa (com talvez centenas de linha no arquivo sob análise). Vamos ter que construir a tabela utilizando algumas funções existentes no Calc ®, como "MÁXIMO", "MÍNIMO", "CONT.SE", entre outras.

2.2.1 - Tabela de frequências agrupada em classes

Vamos observar os passos apresentados em aula:

1) Determinar o intervalo do conjunto

Para tanto temos que identificar os extremos do conjunto, seus valores máximo e mínimo. Vamos começar pelo mínimo. Escolha uma célula onde você deseja que o resultado seja colocado: Selecione esta célula com o cursor. Para inserir uma função basta pressionar “Ctrl F2” (ou através do menu “Inserir” a opção “Função”, e você terá a seguinte caixa de diálogo):

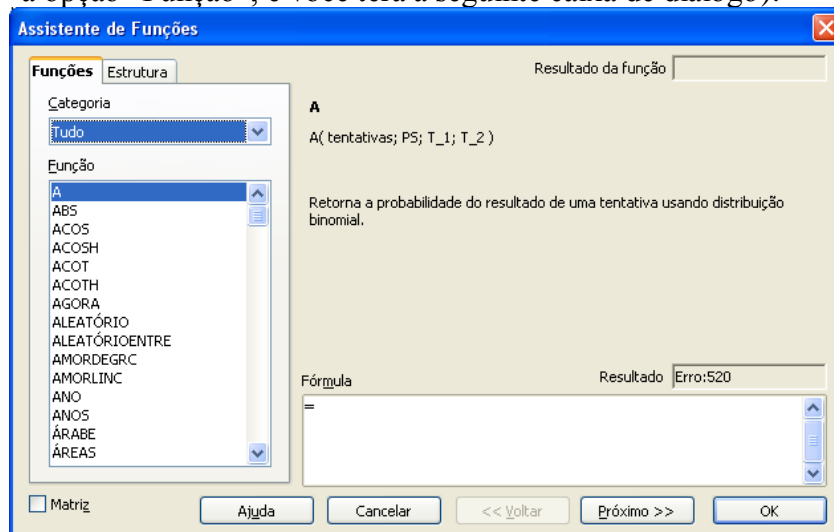


Figura 53 – Funções disponíveis no Calc ®

Você pode procurar diretamente a função desejada no campo Função. Ou, fazer uma seleção mais acurada através do campo categoria.

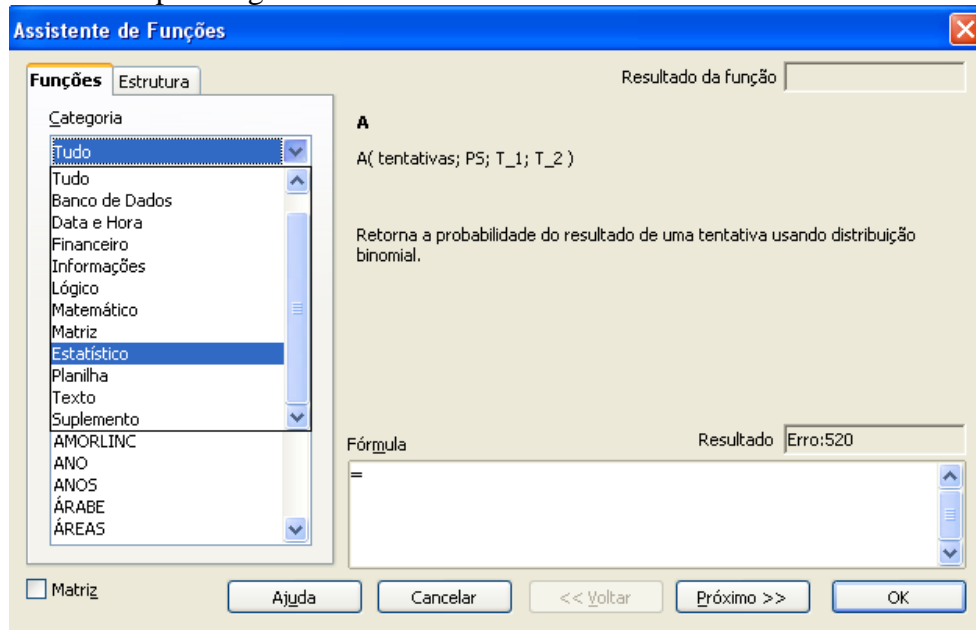


Figura 54 – Funções disponíveis no Calc ® por categoria

Observe que há várias categorias de função incluindo “Estatístico”. Ao selecionarmos esta teremos a seguinte tela:

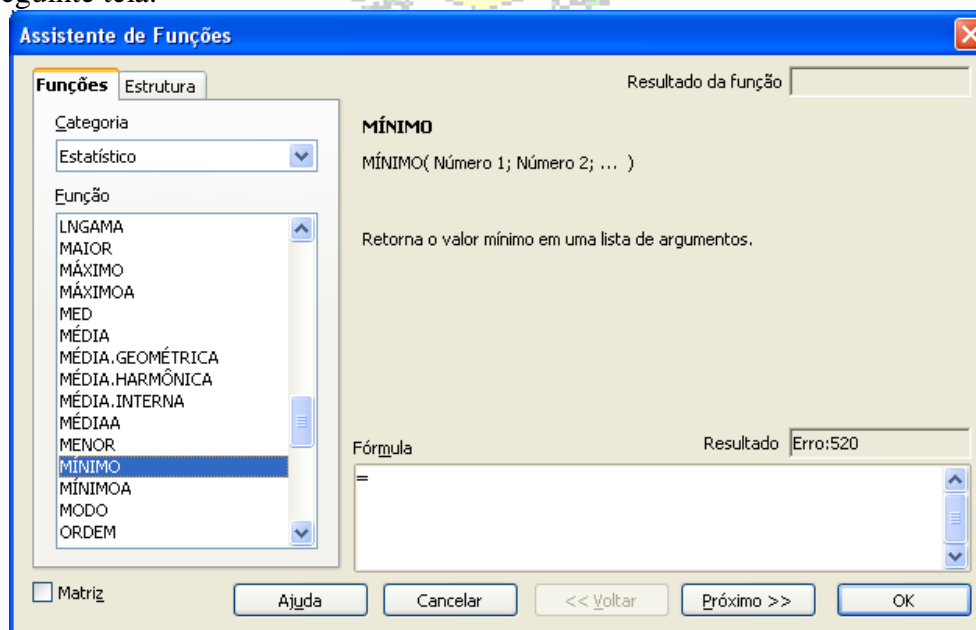


Figura 55 – Funções estatísticas disponíveis no Calc ®

Há uma infinidade de funções estatísticas, observe na figura acima que há uma breve descrição de cada uma delas, incluindo sua sintaxe e argumentos, o que pode facilitar muito a encontrar o que se deseja. No presente problema estamos interessados nas funções "MÁXIMO" e "MÍNIMO". Procurando pela função “MÍNIMO”, e pressionando “Próximo” chegamos à figura 56:

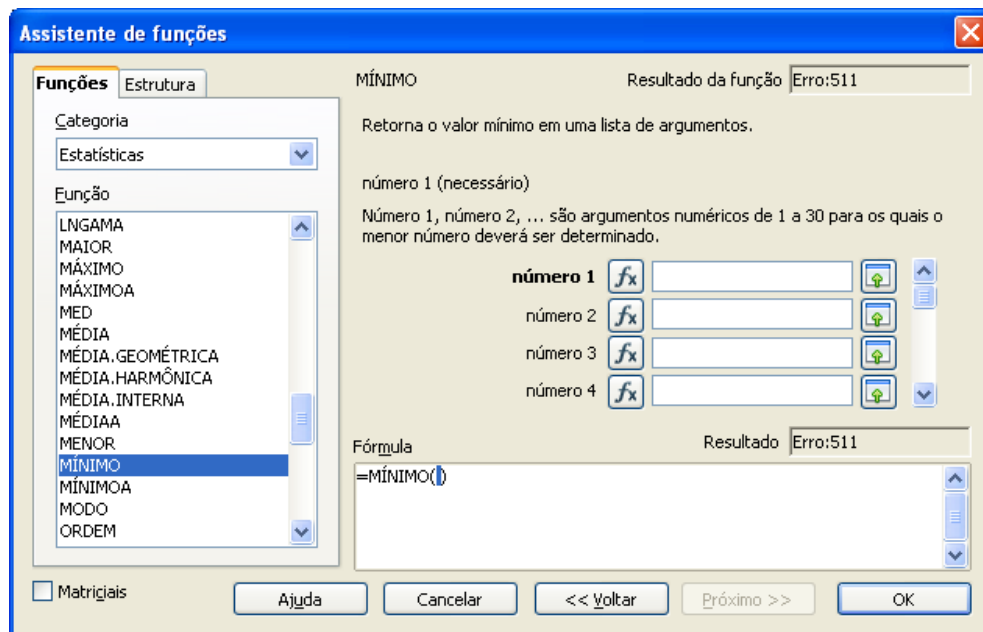


Figura 56 - Função Mínimo

Você precisa escolher o intervalo de valores da variável, para que o Calc possa identificar o valor mínimo: neste caso o intervalo das células G2 a G251, correspondentes aos valores de Renda. Observe o resultado da fórmula, e o link “Ajuda”, que pode ser muito útil.

O mesmo resultado poderia ser obtido simplesmente digitando a fórmula diretamente na célula: $= \text{MÍNIMO}(G2:G251)$ (podem ser usadas maiúsculas ou minúsculas).

Para encontrar o valor máximo podemos realizar um procedimento análogo utilizando a função "MÁXIMO", mas colocando o resultado em outra célula. Para calcular o intervalo podemos colocar uma fórmula em uma célula, fazendo a subtração entre máximo e mínimo.

		N	O
1			
2	Mínimo		1,795
3	Máximo		86,015
4	Intervalo		84,22

Figura 57 - Valores mínimo e máximo de Renda

A menor renda mensal dos clientes pesquisados é de 1,795 salários mínimos (cerca de R\$915,45 em agosto de 2010), e a maior é de 86,015 salários mínimos (cerca R\$ 35 mil em fevereiro de 2009). Este intervalo precisa estar contido na distribuição agrupada em classes. Reparem que o intervalo (diferença entre máximo e mínimo) vale 84,22.

2) Dividir o intervalo em um número conveniente de classes.

Usualmente definimos o número de classe calculando a raiz quadrada do tamanho do conjunto. No arquivo “AmostraToyord.ods” há 250 pessoas, extraindo a raiz quadrada (a função RAIZ, na categoria "Matemáticas" do Calc ®), obtemos 15,81. Poderíamos utilizar 16 classes, mas este valor seria um tanto elevado (criaríamos muitas categorias, o que poderia complicar outras análises que faremos no futuro): usaremos apenas 5 classes, para demonstrar o procedimento. Com isso a amplitude das classes seria igual a 84,22 (intervalo) dividido por 5, resultando que cada classe teria uma amplitude de 16,844. Veja a Figura 58.

	N	O
1		
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5		
6	Classes	5
7		
8	Amplitude	=O4/O6

3) Estabelecer os limites das classes.

Podemos arbitrar valores diferentes para a amplitude das classes e o valor inicial, desde que este seja menor do que 1,795 (1,7, por exemplo), e a amplitude garanta que o valor máximo será incluído no conjunto. Escolhendo uma amplitude de 16,84, e um valor inicial igual ao mínimo do conjunto, tendo em mente que a tabela terá 5 classes, o resultado poderá ser calculado em células apropriadas, como mostrado a partir da Figura 59.

Figura 58 - Amplitude de classes de Renda

	N	O
1		
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5		
6	Classes	5
7		
8	Amplitude	16,84
9		
10	Classes	
11	Limite inferior	Limite superior
12	=O2	1,795

	N	O
1		
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5		
6	Classes	5
7		
8	Amplitude	16,84
9		
10	Classes	
11	Limite inferior	Limite superior
12		=N12+\$O\$8

Figura 59 – Cálculo dos limites inferior e do limite superior da primeira classe de Renda

O limite inferior da primeira classe será o conteúdo da célula O2 (mínimo de Renda), e o limite superior será o limite inferior (célula N12) mais o conteúdo da célula O8 (amplitude das classes). A amplitude das classes é referência absoluta (basta pôr \$ antes de cada letra e número do intervalo que estiver sendo usado), para possibilitar arrastar a fórmula até o final da tabela. Precisamos também definir os limites da segunda classe da tabela, na Figura 60.

	N	O
1		
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5		
6	Classes	5
7		
8	Amplitude	16,84
9		
10	Classes	
11	Limite inferior	Limite superior
12	1,795	18,64
13	=O12	

	N	O
1		
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5		
6	Classes	5
7		
8	Amplitude	16,84
9		
10	Classes	
11	Limite inferior	Limite superior
12	1,795	18,64
13		=N13+\$O\$8

Figura 60 – Cálculo dos limites inferior e superior da segunda classe de Renda

O limite inferior da segunda classe é o limite *superior* da primeira classe. E o limite superior é igual ao limite inferior mais amplitude das classes. Agora sim, podemos arrastar as células N13 e O13 até o fim da tabela e completar as 5 classes. Veja a Figura 61.

	N	O
1		
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5		
6	Classes	5
7		
8	Amplitude	16,84
9		
10	Classes	
11	Limite inferior	Limite superior
12	1,795	18,639
13	18,639	35,483
14	35,483	52,327
15	52,327	69,171
16	69,171	86,015

4) Determinar as frequências de cada classe é o passo mais difícil. Podemos resolver o problema se utilizarmos a função "CONT.SE". Esta função conta quantos valores em um determinado intervalo de dados atendem a um critério estabelecido.

Podemos montar uma fórmula para obter a quantidade de observações entre o limite inferior e superior de uma classe. Apenas por conveniência, costuma-se supor que o limite superior apenas limita a classe: se, por exemplo, houver uma renda de 9,578 salários mínimos, ela será registrada na terceira classe e não na segunda. Apenas o limite superior da *última* classe, caso ele coincidir com o valor máximo da variável, será considerado como *dentro* da classe.

Precisamos também definir os pontos médios das classes, que passarão a representá-las, como se as frequências obtidas fossem deles.

Esses procedimentos são mostrados nas Figuras 61 a 63.

Figura 61 - Classes de Renda

	N	O	P	Q	R
1					
2	Mínimo	1,795			
3	Máximo	86,015			
4	Intervalo	84,22			
5					
6	Classes	5			
7					
8	Amplitude	16,84			
9					
10	Classes				
11	Limite inferior	Limite superior	Ponto médio	Frequência	%
12	1,795	18,639	$=(N12+O12)/2$		
13	18,639	35,483			
14	35,483	52,327			
15	52,327	69,171			
16	69,171	86,015			

Figura 62 - Ponto médio da primeira classe de Renda

Basta arrastar a fórmula da célula P12 até a P16 para obter os pontos médios das classes.

	N	O	P	Q	R	S	T	U	V
1									
2	Mínimo	1,795							
3	Máximo	86,015							
4	Intervalo	84,22							
5									
6	Classes	5							
7									
8	Amplitude	16,84							
9									
10	Classes								
11	Limite inferior	Limite superior	Ponto médio	Freqüência	%				
12	1,795	18,639	10,22	=CONT.SE(\$G\$2:\$G\$251;"<"&O12)-CONT.SE(\$G\$2:\$G\$251;"<"&N12)					
13	18,639	35,483							
14	35,483	52,327							
15	52,327	69,171							
16	69,171	86,015							

Figura 63 - Obtenção da Freqüência da primeira classe de Renda

Observe que contam-se as ocorrências menores do que o limite superior da classe, subtraídas da contagem das ocorrências menores do que o limite inferior, mantendo o intervalo G2 a G251 (valores de Renda) como referência absoluta, para permitir arrastar a fórmula até a célula Q16.

	N	O	P	Q	R	S	T	U	V
1									
2	Mínimo	1,795							
3	Máximo	86,015							
4	Intervalo	84,22							
5									
6	Classes	5							
7									
8	Amplitude	16,84							
9									
10	Classes								
11	Limite inferior	Limite superior	Ponto médio	Freqüência	%				
12	1,795	18,639	10,22	98					
13	18,639	35,483	27,06	102					
14	35,483	52,327	43,91	38					
15	52,327	69,171	60,75	9					
16	69,171	86,015	77,59	=CONT.SE(\$G\$2:\$G\$251;"<="&O16)-CONT.SE(\$G\$2:\$G\$251;"<"&N16)					

Figura 64 - Obtenção da Freqüência da última classe de Renda

Observe que na obtenção da Freqüência da última classe de Renda devemos usar o critério \leq quando lidando com o limite superior da classe, pois ele coincide com o valor máximo do conjunto. Se não fizermos isso, a contagem desconsiderará este valor máximo, e a tabela ficará incorreta.

Podemos também obter freqüências e percentuais acumulados para a tabela de Renda. Basta

remover o segundo membro das fórmulas mostradas nas Figuras 63 e 64, tomando os mesmos cuidados com o limite superior da última classe, como mostrado nas Figuras 65 e 66.

	N	O	P	Q	R	S	T	U
1								
2	Mínimo	1,795						
3	Máximo	86,015						
4	Intervalo	84,22						
5								
6	Classes	5						
7								
8	Amplitude	16,84						
9								
10	Classes							
11	Limite inferior	Limite superior	Ponto médio	Freqüência	%	Freq. acumulada	% acumulado	
12	1,795	18,639	10,22	98	39,20%	=CONT.SE(\$G\$2:\$G\$251;"<"&O12)		
13	18,639	35,483	27,06	102	40,80%			
14	35,483	52,327	43,91	38	15,20%			
15	52,327	69,171	60,75	9	3,60%			
16	69,171	86,015	77,59	3	1,20%			
17			Total	250	100,00%			

Figura 65 - Obtenção da Freqüência acumulada da primeira classe de Renda

	N	O	P	Q	R	S	T	U
1								
2	Mínimo	1,795						
3	Máximo	86,015						
4	Intervalo	84,22						
5								
6	Classes	5						
7								
8	Amplitude	16,84						
9								
10	Classes							
11	Limite inferior	Limite superior	Ponto médio	Freqüência	%	Freq. acumulada	% acumulado	
12	1,795	18,639	10,22	98	39,20%	98		
13	18,639	35,483	27,06	102	40,80%	200		
14	35,483	52,327	43,91	38	15,20%	238		
15	52,327	69,171	60,75	9	3,60%	247		
16	69,171	86,015	77,59	3	1,20%	=CONT.SE(\$G\$2:\$G\$251;"<="&O16)		
17			Total	250	100,00%			

Figura 66 - Obtenção da Freqüência acumulada da primeira classe de Renda

Após todas as operações, incluindo a obtenção dos percentuais a tabela agrupada em classes da Renda resulta na Figura 67.

	N	O	P	Q	R	S	T
1							
2	Mínimo	1,795					
3	Máximo	86,015					
4	Intervalo	84,22					
5							
6	Classes	5					
7							
8	Amplitude	16,84					
9							
10	Classes						
11	Limite inferior	Limite superior	Ponto médio	Frequência	%	Freq.acumulada	% acumulado
12	1,795	18,639	10,22	98	39,20%	98	39,20%
13	18,639	35,483	27,06	102	40,80%	200	80,00%
14	35,483	52,327	43,91	38	15,20%	238	95,20%
15	52,327	69,171	60,75	9	3,60%	247	98,80%
16	69,171	86,015	77,59	3	1,20%	250	100,00%
17			Total	250	100,00%		

Figura 67 - Renda agrupada em classes: frequências absolutas, percentuais, frequências acumuladas e percentuais acumulados

Há maior número de clientes com renda mais “baixa” (até 35,495 salários mínimos, cerca de R\$ 18102 em agosto de 2010, nem tão baixa assim...¹). Esta tabela pode ser usada para construir um histograma, de forma semelhante ao que foi visto no item 2.1. Os dados serão as células com as frequências (incluindo a célula com o nome “Frequências”) e no campo “Categorias” podemos colocar as células com os pontos médios das classes. Seleciona-se o gráfico, surgindo a tela da figura 13. Escolhemos gráfico em colunas 2D e pressionamos “Próximo”, chegando à figura 68.

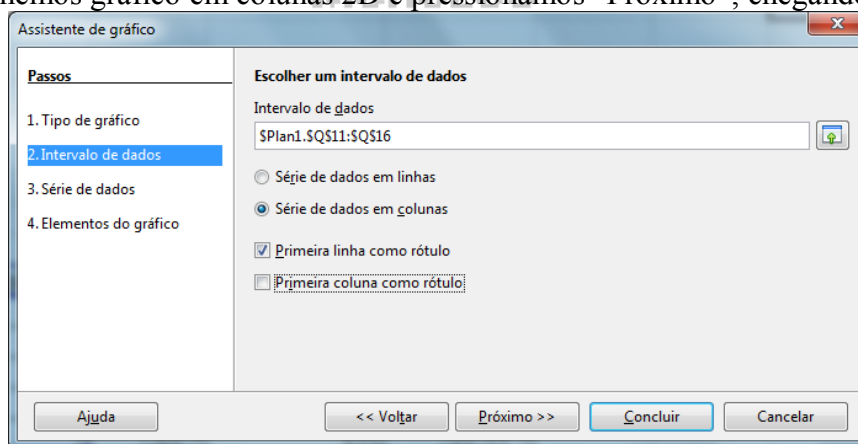


Figura 68 – Assistente gráfico 2ª etapa

Alguns detalhes *muito* importantes:

- os dados estão nas células Q11 a Q16 (o nome Frequências, e as frequências das 5 classes);
- os dados estão em uma coluna, então é preciso marcar “Série de dados em colunas”;
- a primeira linha do intervalo contém o rótulo dos dados (Frequência, no caso), então marca-se “Primeira linha como rótulo”;
- *desmarca-se* “Primeira coluna como rótulo”, para evitar problemas.

Pressionando “Próximo” podemos acrescentar as categorias que aparecerão no eixo horizontal do gráfico, figura 69.

¹ Talvez fosse mais apropriado ter usado 10 classes.

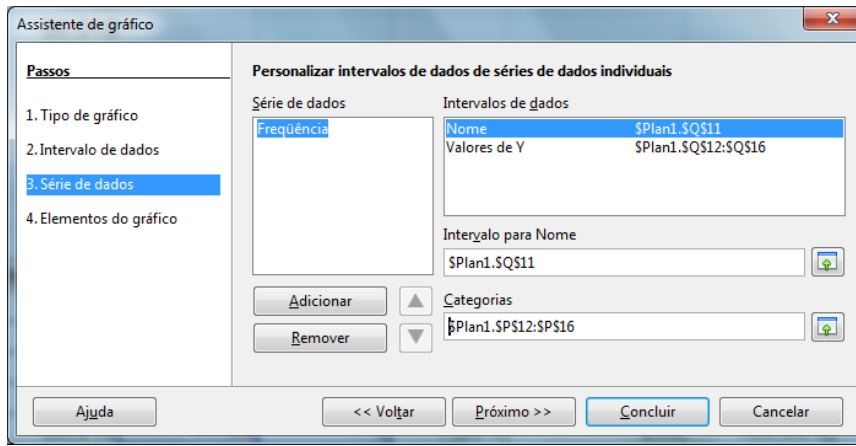


Figura 69 – Assistente gráfico 3ª etapa

No campo “Categorias” podemos incluir as células P12 a P16 que contêm os pontos médios das classes. Ou, se quiséssemos, poderíamos escrever os limites das classes em células específicas e referenciá-las no campo. Pressionar “Próximo”, chegando à figura 70.

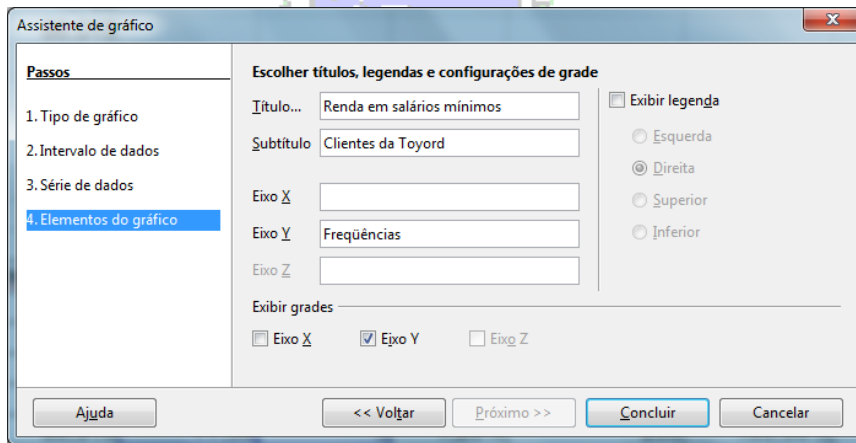


Figura 70 – Assistente gráfico 3ª etapa

Não há necessidade de exibir legenda, então se desmarca esta opção. Preenche-se o título, subtítulo e título do eixo Y. Nas figuras 48 a 51 (item 2.1) explicou-se como eliminar o espaçamento entre as barras para criar um histograma: se fizermos isso com o gráfico de renda chegamos finalmente à figura 71.

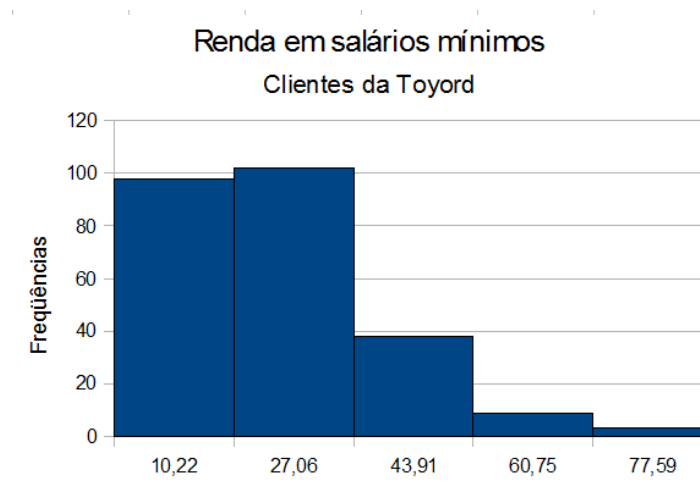


Figura 71 - Histograma de Renda (agrupado em classes)

2.2 - Categorização de uma variável quantitativa contínua

Em algumas situações é interessante transformar uma variável quantitativa contínua em qualitativa para realizar certas análises. Por exemplo, em uma pesquisa de mercado poderia haver interesse em fazer a análise conjunta entre a opinião sobre um produto e a renda mensal do entrevistado.

Para o caso da variável Renda mensal dos clientes da montadora, no arquivo “AmostraToyord.ods”, podemos adotar a seguinte categorização (totalmente arbitrária): os clientes com renda menor do que 4 salários mínimos serão classificados como de renda baixa; aqueles com renda maior ou igual a 4 e menor do que 20 salários mínimos serão de renda média; e os restantes, com renda maior ou igual a 20 salários mínimos serão de renda alta.

Precisamos criar uma nova variável em uma coluna livre da planilha do Calc ®. Para facilitar o processo, porém, é ideal que você selecione com o cursor uma célula na mesma linha do primeiro valor de renda, que está na linha 2: por exemplo, no arquivo “AmostraToyord.ods”, você poderia pôr o cursor na célula Z1. Digite o nome da nova variável, “Renda Categorizada”, e passe o cursor para a célula Z2. A variável "Renda Categorizada" deverá relacionar-se com a variável "Renda", transformando os valores numéricos em categorias. Isto pode ser feito através de uma função do Calc ®, uma função lógica, a função SE. Podemos inseri-la na célula Z2, logo abaixo do nome da nova variável: escolha a categoria "Lógica", e dentro desta a função SE.

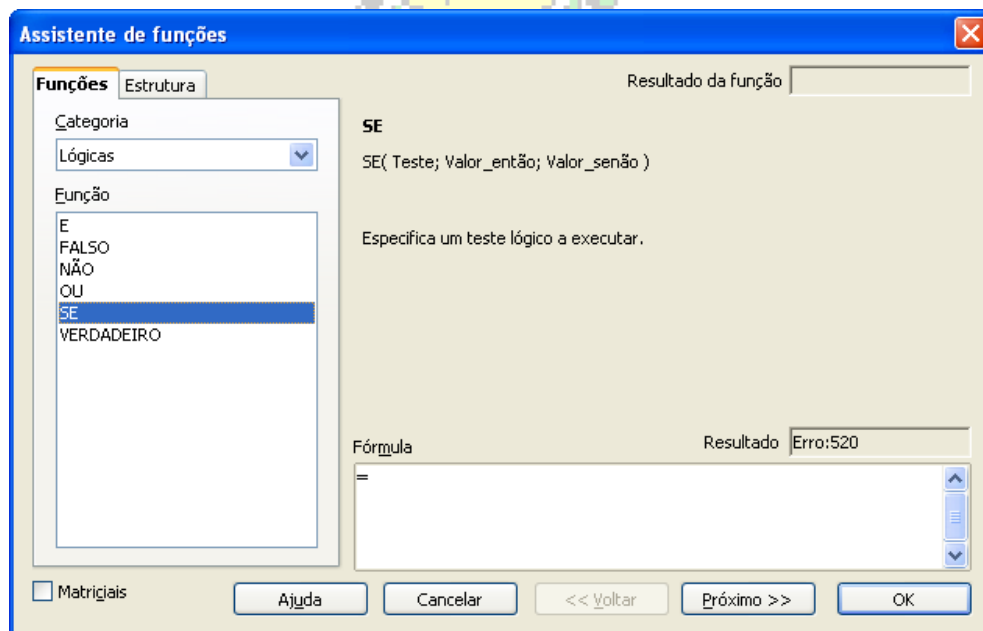


Figura 72 - Função lógica SE

Esta função permitirá categorizar a variável "Renda", registrando seus valores na variável "Renda Categorizada". O teste lógico será realizado através da comparação dos valores de "Renda" com os limites das categorias de Renda, resultando em diferentes valores que serão registrados na variável "Renda Categorizada". Pressionando "OK" surgirá a tela apresentada na figura 73.

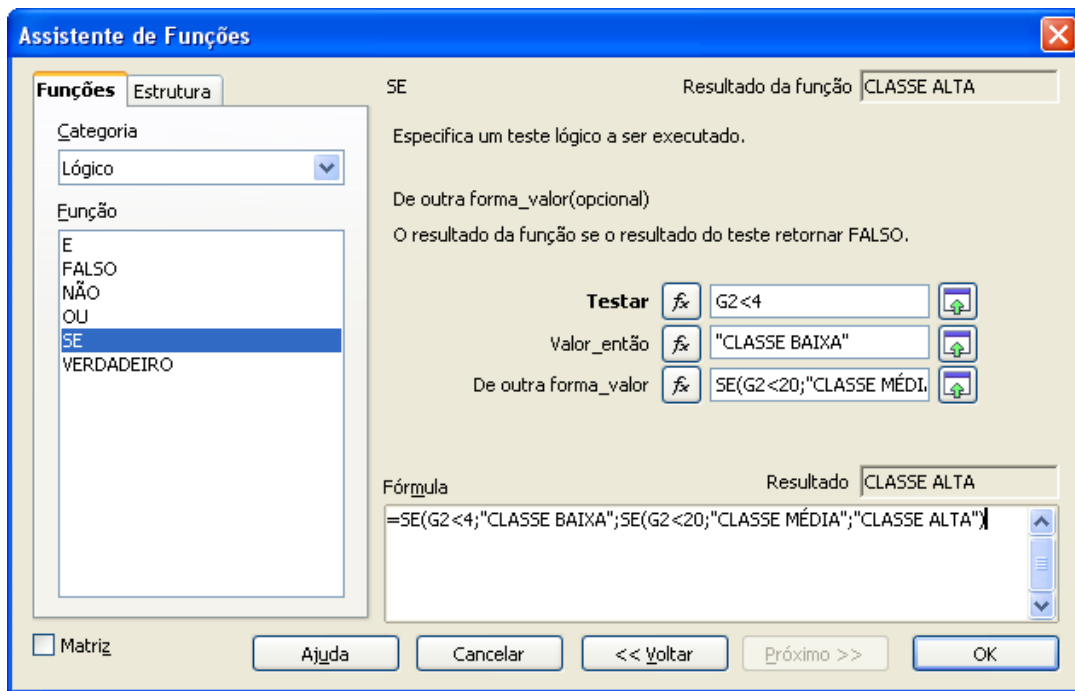


Figura 73 - Argumentos da função SE

Observe que na opção “De outra forma_valor” (ou Valor_senão em algumas versões do Calc ®) podemos pôr outra função SE para permitir fazer mais comparações. Neste caso, se a renda for menor do que 4 o indivíduo será considerado “classe baixa” (as aspas são obrigatórias para a categorização). Caso não seja, a renda será comparada com 20 (já se sabe que é maior do que 4), caso seja menor o indivíduo será considerado “classe média”. Se não for menor do que 20 resta apenas a opção de classificá-lo como “classe alta”.

A estrutura mostrada na figura 73 pode ser representada de outra forma, escrevendo diretamente a fórmula na célula.

```
=SE(G2<4;"Classe baixa";SE(G2<20;"Classe média";"Classe alta"))
```

Observe que os valores precisam estar entre aspas. Como temos que realizar vários testes, na mesma expressão, para avaliar a qual das classes pertence o valor, deveremos inserir outras funções SE nos espaços referentes aos valores que a célula teria se o teste lógico desse resultado falso. Observe a estrutura a seguir, registrada em um fluxograma, mostrado na figura 74.

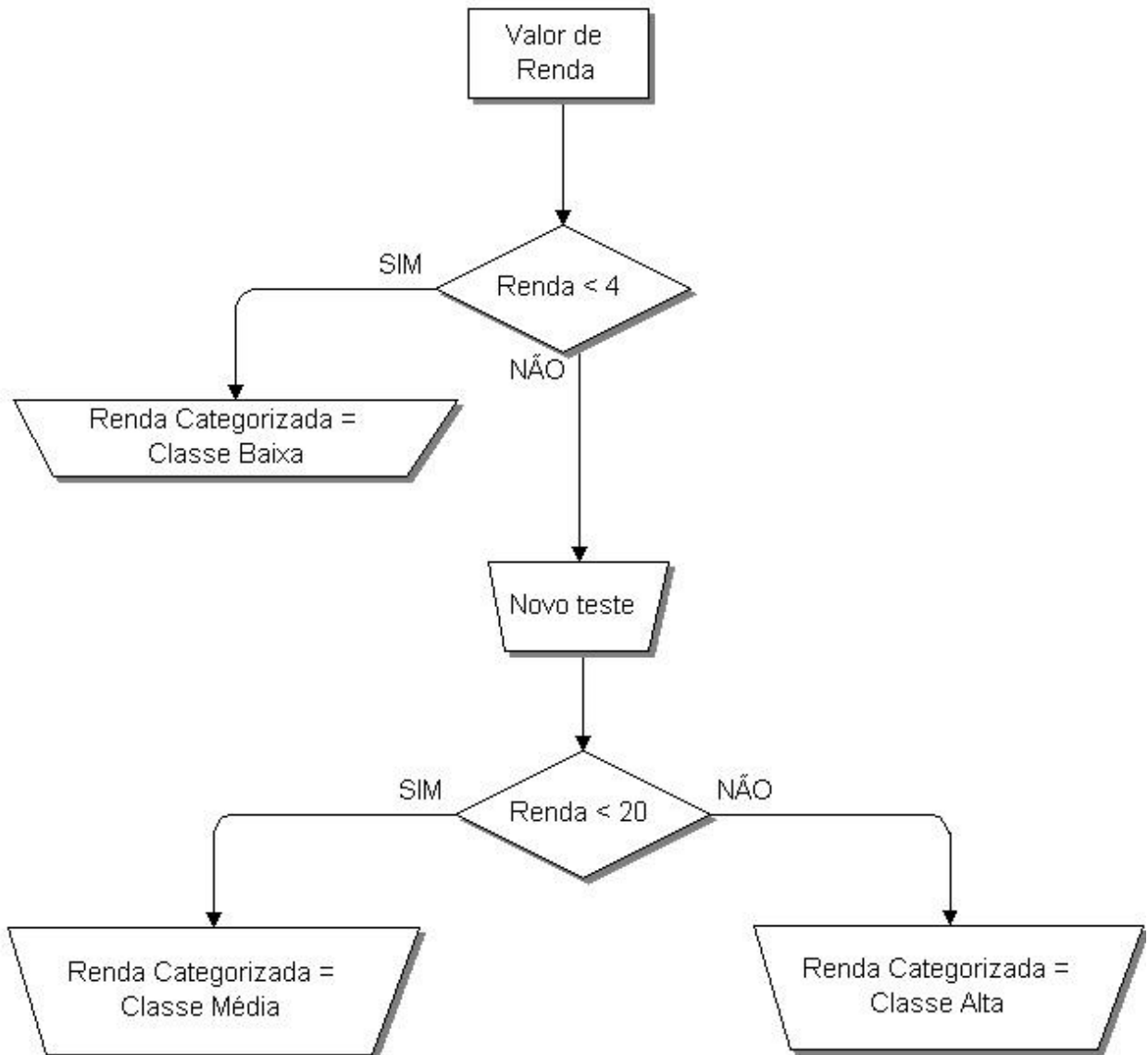


Figura 74 - Procedimento para categorizar a variável Renda

Observe com cuidado a estrutura da função:

- 1) Testa-se se a renda é menor do que 4. Se for, a renda será categorizada como classe baixa, senão novo teste será realizado.
- 2) Já se sabe que a renda é maior ou igual a 4, então basta testar se é menor do que 20. Se for, a renda será categorizada como classe média, senão será classe alta, já que não há mais nenhuma opção possível.

Uma vez completa a fórmula na célula Z2 podemos "arrastá-la" para baixo, chegando ao final do arquivo, na célula Z251. O resultado será uma nova variável, relacionada à Renda, mas categorizada (qualitativa), veja a figura 70.

Z2		f(x)	Σ	=	=SE(G2<4;"CLASSE BAIXA";SE(G2<20;"CLASSE MÉDIA";"CLASSE ALTA"))				
	Z	AA	AB	AC	AD	AE	AF	AG	
1	Renda categorizada								
2	CLASSE ALTA								
3	CLASSE ALTA								
4	CLASSE ALTA								
5	CLASSE MÉDIA								
6	CLASSE MÉDIA								
7	CLASSE ALTA								

Figura 75 - Variável Renda Categorizada
Fonte: adaptada pelo autor de Sun Microsystems ®

Utilizando um procedimento semelhante ao visto no item 1.1 para a variável Renda Categorizada:

Renda categorizada		%
CLASSE ALTA	144	57,60
CLASSE BAIXA	2	0,80
CLASSE MÉDIA	104	41,60
Total Resultado	250	100,00

Figura 76 - Tabela dinâmica da variável Renda Categorizada
Fonte: adaptada pelo autor de Sun Microsystems ®

Talvez a categorização escolhida não tenha sido a melhor... Observe que na Classe baixa há apenas duas observações.

Podemos agora fazer cruzamentos com variáveis qualitativas, através de tabelas dinâmicas, através do procedimento visto no item 1.2. Também podemos construir gráficos para variáveis qualitativas (de barras ou em setores).

2.3 - Cálculo de Medidas de Síntese

Ao lidar com variáveis quantitativas geralmente há interesse em calcular medidas de síntese, para obter mais informações sobre as variáveis sob análise (veja a Unidade 4). O Calc ® permite calcular uma série de medidas através de funções pré-programadas. Isso pode ser feito diretamente, inserindo as funções, ou selecionando-as através de uma tabela dinâmica.

2.3.1 - Cálculo de Medidas de Síntese através de inserção de funções

Vamos relacionar as principais medidas de síntese com as funções do Calc.

Quadro 1 - Medidas de síntese e funções do Calc

MEDIDAS DE SÍNTESE	FUNÇÕES DO calc
Média aritmética simples (média)	MÉDIA(intervalo de dados)
Mediana	MED(intervalo de dados)
Moda	MODO(intervalo de dados)
Mínimo	MÍNIMO(intervalo de dados)
Máximo	MÁXIMO(intervalo de dados)
Desvio padrão amostral (n-1)	DESVPAD(intervalo de dados)
Desvio padrão populacional (n)	DESVPADP(intervalo de dados)
Quartil inferior (Q1)	QUARTIL(intervalo de dados;1)
Quartil superior (Q3)	QUARTIL(intervalo de dados;3)

Fonte: elaborado pelo autor

Para calcular as medidas de síntese basta inserir as fórmulas nas células de interesse, selecionar os intervalos de dados, e pressionar "ENTER" e o Calc realizará o cálculo. Para os dados presentes no arquivo "AmostraToyord.ods", imagine que quiséssemos calcular as seguintes medidas de síntese para a variável "Renda": média, mediana, moda, mínimo, máximo, desvio padrão, quartil inferior e quartil superior (oito medidas ao todo). Vamos calcular o desvio padrão amostral pois os dados referem-se a uma amostra dos clientes da montadora. Sabemos que os dados de "Renda" estão no intervalo das células G2 a G251. Primeiramente escrevemos os nomes das medidas, em quaisquer células livres da planilha Depois começamos a inserir as funções ao lado dos nomes, indicando o intervalo de dados:

=MÉDIA(G2:G251)
 =MED(G2:G251)
 =MODO(G2:G251)

=MÍNIMO(G2:G251)
 =MÁXIMO(G2:G251)
 =DESVPAD(G2:G251)
 =QUARTIL(G2:G251;1)
 =QUARTIL(G2:G251;3)

Os resultados serão:

Média	25,06216
Mediana	23,405
Moda	6,645
Mínimo	1,795
Máximo	86,015
Desvio padrão	14,4922
Quartil Inferior	14,11875
Quartil Superior	32,1275

Figura 77 - Medidas de síntese da variável Renda
 Fonte: adaptada pelo autor de Sun Microsystems ®

Observe que os resultados obtidos são para a totalidade dos salários, ou seja, não há segmentação em função de outras variáveis, o que poderia ser interessante para a análise. A utilização de funções só permitiria a obtenção de medidas segmentadas por outras variáveis se a base de dados for rearranjada, de maneira a agrupar os dados em função dos valores da variável "independente". Para obter medidas segmentadas, ou seja realizar uma análise "breakdown" precisamos usar as tabelas dinâmicas. O problema com o Calc é a impossibilidade de pôr mais de uma medida na mesma tabela, assim, é preciso usar um expediente diferente a classificação da base de dados em função da variável independente, e a partir da ordenação resultante calcular as medidas (este procedimento também pode ser feito com o Excel ®).

2.3.2 - Cálculo de medidas de síntese utilizando classificação em função de uma variável independente.

No item 1.1, figura 5, pudemos observar que há uma série de medidas que podem ser calculadas através de uma tabela dinâmica. A tabela dinâmica permite identificar os resultados em função de uma ou mais variáveis independentes, o que permitirá uma análise mais pormenorizada dos dados. Mas no Calc ® (ao contrário do Excel ®) não é possível pôr mais de uma medida na tabela dinâmica: cada medida precisa de uma tabela própria, o que torna o processo tedioso e a interpretação dos resultados mais difícil. Torna-se mais interessante classificar a base de dados em função da variável independente (qualitativa ou categorizada), e, observando a ordenação, calcular as medidas para os intervalos da variável dependente (quantitativa) correspondentes a cada categoria da independente.

Vamos imaginar que há interesse em avaliar como está a distribuição da variável "Renda" em função da variável "Modelo", no arquivo "AmostraToyord.ods". É extremamente interessante conhecer o perfil econômico dos clientes por modelo. Então vamos reordenar toda a base de dados em função da variável Modelo.

Devemos seleccionar a primeira célula da base de dados (A1). Em seguida, no menu Dados escolhemos a opção classificar:

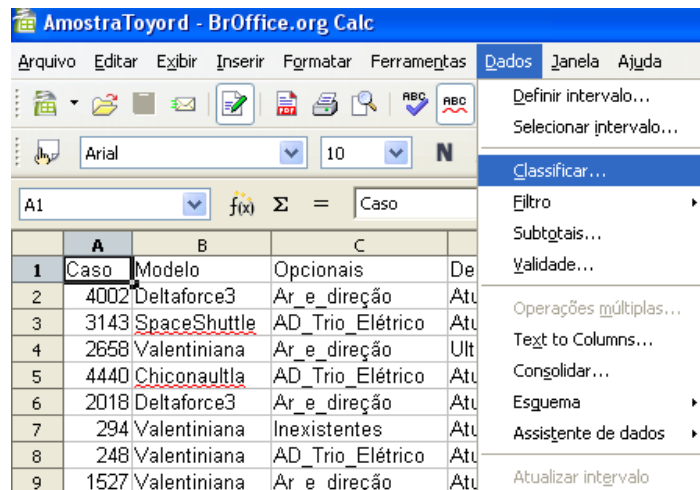


Figura 78 – Menu Dados: opção Classificar
 Fonte: adaptada pelo autor de Sun Microsystems ®

O resultado da seleção será a caixa de diálogo mostrada na figura 74:

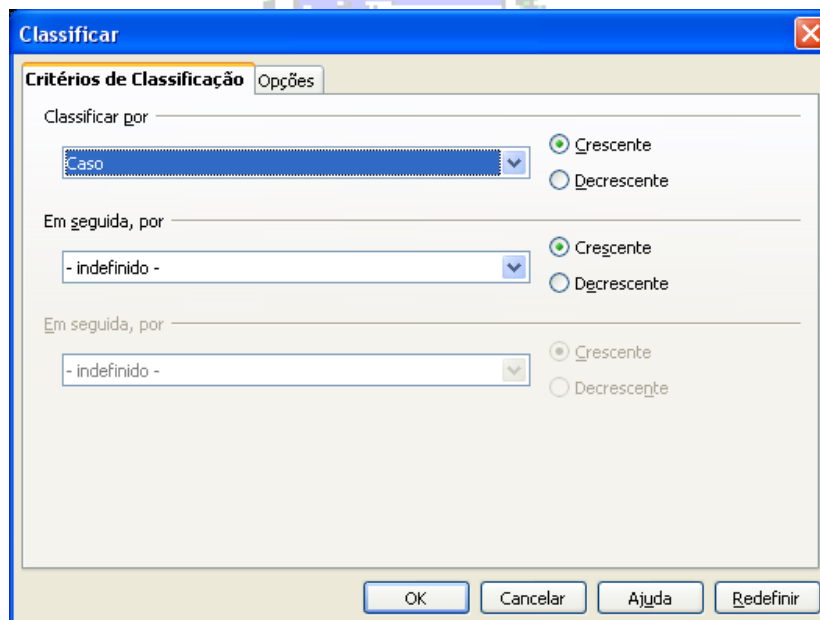


Figura 79 – Critérios de classificação
 Fonte: adaptada pelo autor de Sun Microsystems ®

Automaticamente o Calc ® seleciona todas as variáveis da base de dados e permite que ordenemos a base de dados em função de até três variáveis, de forma crescente ou decrescente. Como “Caso” é a primeira variável da base ela é selecionada (padrão), mas podemos selecionar qualquer outra, como visto a seguir:

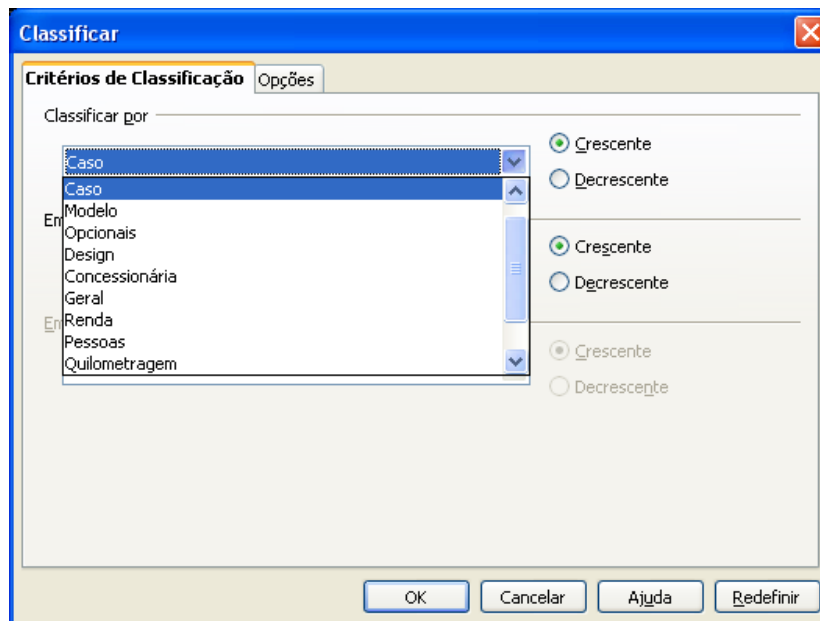


Figura 80 – Critérios de classificação – escolha de variáveis
Fonte: adaptada pelo autor de Sun Microsystems ®

Devemos escolher Modelo e pressionar OK. Automaticamente o Calc ® reordena toda a base de dados em função de Modelo, mas sem “desorganizá-la”, mantendo a coerência das informações. Veja a figura 76 abaixo:

	A	B	C	D	E	F	G
1	Caso	Modelo	Opcionais	Design	Concessionária	Geral	Renda
2	484		Ar_e_direção	Atualizados	Insatisfatória	Insatisfeito	14,29
3	1661	<u>Chiconaultla</u>	Inexistentes	Atualizados	Insatisfatória	Muito insatisfeito	15,51
4	1400	<u>Chiconaultla</u>	Inexistentes	Atualizados	Insatisfatória	Muito insatisfeito	12,6
5	4121	<u>Chiconaultla</u>	Inexistentes	Atualizados	Não causou impressão	Muito insatisfeito	14,99
6	4635	<u>Chiconaultla</u>	Inexistentes	Atualizados	Insatisfatória	Muito insatisfeito	9,16
7	740	<u>Chiconaultla</u>	Ar_e_direção	Adiante dos outros	Muito insatisfatória	Insatisfeito	12,51
8	2963	<u>Chiconaultla</u>	Inexistentes	Adiante dos outros	Muito insatisfatória	Muito insatisfeito	10,5

Figura 81 – Base de dados ordenada em função de Modelo
Fonte: adaptada pelo autor de Sun Microsystems ®

As células vazias são consideradas os “menores” valores da variável, e os demais são organizados de forma crescente. Ao realizar a classificação observamos a seguinte disposição:

- modelo Chiconaultla – linhas 2 a 83.
- modelo DeltaForce3 – linhas 84 a 139.
- modelo LuxuriousCar – linhas 140 a 168.
- modelo SpaceShuttle – linhas 169 a 210.
- modelo Valentiniana – linhas 211 a 251.

Com as informações acima podemos calcular as medidas de síntese de forma segmentada para a variável renda: o intervalo G2:G83 contém as rendas dos proprietários do Chiconaultla, G84:G138 do DeltaForce3, G139:G168 do LuxuriousCar, G169:G210 do SpaceShuttle e G211:G251 do Valentiniana. Basta realizar os cálculos das medidas (ver quadro 1) para cada intervalo, embora possa ser extenso caso a variável independente tenha vários valores, proporciona um conjunto detalhado de medidas. Veja a figura 77.

	AC	AD	AE	AF	AG	AH		
1	Medidas	Chiconaultla	DeltaForce3	LuxuriousCar	SpaceShuttle	Valentiniana		
2	Média	12,7046	22,0636	50,9319	33,0504	27,3532	Média	25,06216
3	Mediana	12,2450	21,3775	47,5250	33,8500	25,7150	Mediana	23,405
4	Modo	6,6450	14,0950	58,9200	25,2700	24,8350	Moda	6,645
5	Mínimo	1,7950	10,8200	29,8000	18,8650	13,0550	Mínimo	1,795
6	Máximo	40,1600	48,2200	86,0150	47,3000	65,3900	Máximo	86,015
7	Desvio padrão	6,0379	6,9661	14,9220	7,6195	8,3827	Desvio padrão	14,4922
8	Quartil inferior	8,8800	16,6850	41,8900	26,6588	23,6850	Quartil Inferior	14,11875
9	Quartil superior	15,4000	26,2788	58,9200	39,3525	30,1300	Quartil Superior	32,1275

Figura 82 – Medidas de síntese de Renda em função de Modelo; medidas gerais de Renda

Fonte: adaptada pelo autor de Sun Microsystems ®

Há clara relação entre Renda e Modelo de veículo. Os proprietários de LuxuriousCar têm a maior renda mínima (29,8 salários mínimos), maior média de renda (50,93 salários mínimos), maior mediana de renda (47,52 salários mínimos), maiores quartis (41,89 e 58,92 salários mínimos) e a maior renda de todo o conjunto (86,015 salários mínimos). Por outro lado, os compradores de Chiconaultla e DeltaForce3 têm as menores medidas de renda.

Observe à direita da figura 77 as medidas para o conjunto como um todo, exatamente iguais às mostradas na figura 72. Percebe-se claramente que os proprietários de LuxuriousCar e SpaceShuttle têm rendas superiores à média geral e os de Chiconaultla e DeltaForce3 inferiores. Já os proprietários de Valentiniana apresentam medidas de síntese próximas as do total do conjunto.

2.3 – Procedimentos para duas variáveis quantitativas (diagrama de dispersão)

Imagine que houvesse interesse em avaliar se clientes com maior renda rodam mais com seus veículos. Os dados estão no arquivo “AmostraToyord.ods”, contém as informações sobre 250 pares de observações Renda (em salários mínimos) e Quilometragem mensal (em km). É razoável imaginar que Renda possa influenciar Quilometragem: com mais dinheiro pode-se gastar mais com combustível e rodar mais com o veículo.

O primeiro passo é construir um diagrama de dispersão relacionando as duas variáveis: Renda será a variável independente (posta no eixo X) e Quilometragem a dependente (no eixo Y).

Clicando sobre o ícone "Assistente Gráfico", na barra de ferramentas do Calc, resultando na figura 13. Selecionando o gráfico Dispersão (XY), obtemos a figura 78.

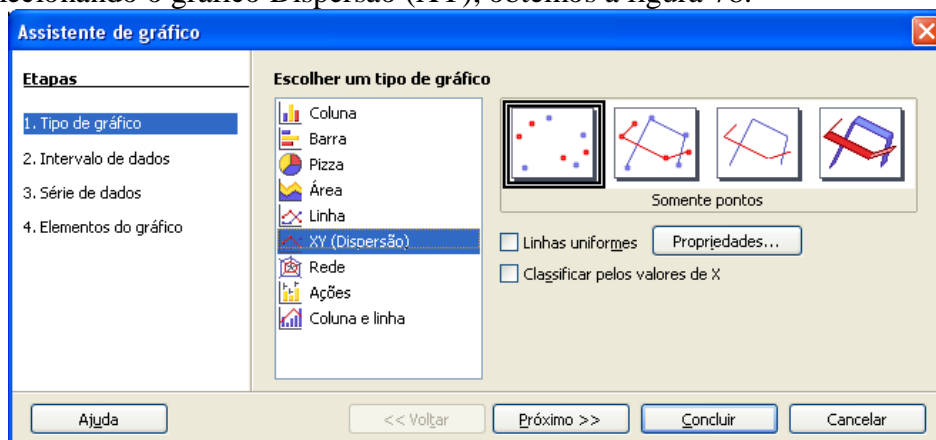


Figura 83 – Assistente gráfico do Calc – diagrama de dispersão – 1ª parte

Fonte: adaptada pelo autor de Sun Microsystems ®

Para os nossos interesses o subtipo mais interessante é o padrão, marcado em preto na figura 78. Pressionando "Próximo" chegaremos a uma tela semelhante à figura 79.

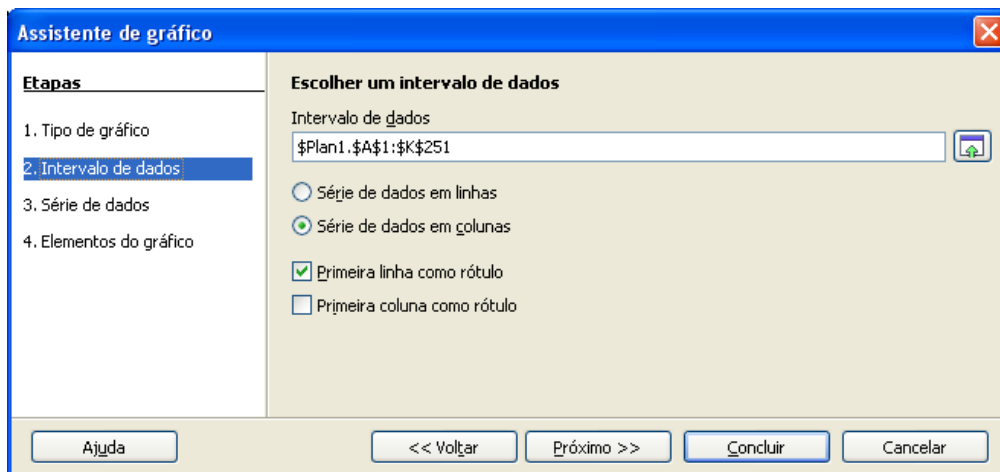


Figura 84 – Assistente gráfico do Calc – diagrama de dispersão – 2ª parte

Fonte: adaptada pelo autor de Sun Microsystems ®

O Calc ® selecionou toda a base de dados (intervalo A1:K251). Caso isso não ocorra o intervalo pode ser selecionado manualmente. Pressionando “Próximo” vamos à figura 80.

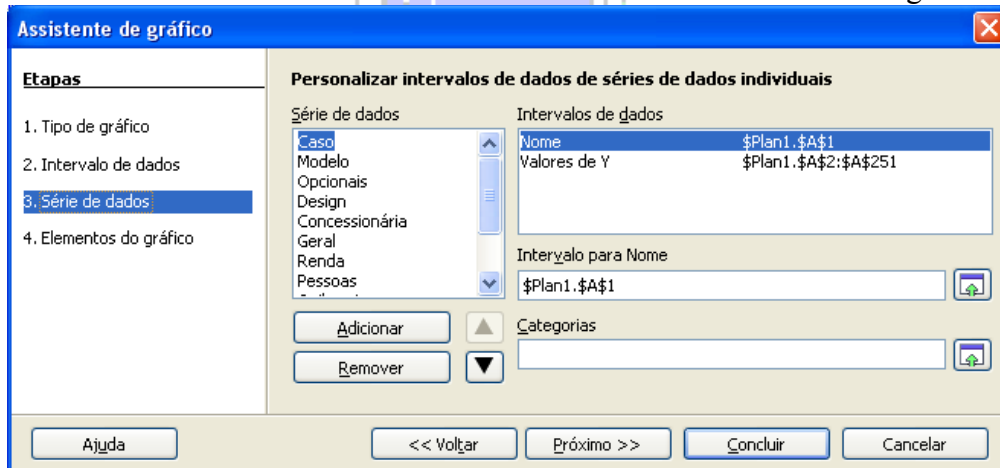


Figura 85 – Assistente gráfico do Calc – diagrama de dispersão – 3ª parte

Fonte: adaptada pelo autor de Sun Microsystems ®

Por ter selecionado toda a base de dados o Calc ® pôs todas as variáveis na série de dados. Ele tentará construir diagramas de dispersão para cada grupo de duas variáveis. Nós não queremos isso, apenas temos interesse em Renda x Quilometragem. Mas, para colocá-la corretamente no gráfico precisamos remover todas as séries existentes e adicionar uma nova (Série sem nome), o que nos leva à figura 81.

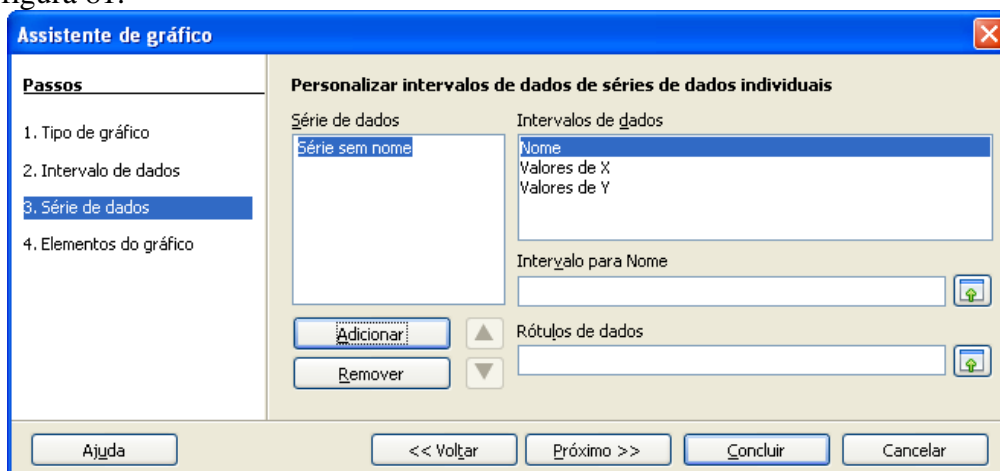


Figura 86 – Assistente gráfico do Calc – diagrama de dispersão – 3ª parte - seleção

Fonte: adaptada pelo autor de Sun Microsystems ®

Basta selecionar “Valores de X” no campo Intervalo de Dados, e no campo Intervalo para nome colocamos o intervalo dos valores de renda (G2:G251). Marque as células na planilha (pressionando a seta verde na extrema direita de cada janela, e marcando as células de interesse na planilha), não digite diretamente, pois causará erro. Em seguida selecionamos “Valores de Y” e colocamos no Intervalo para nome os valores de quilometragem (I2:I251), como na figura 82 (o Calc passa a chamar a Série de Dados de Coluna I, pois são os últimos dados inseridos):

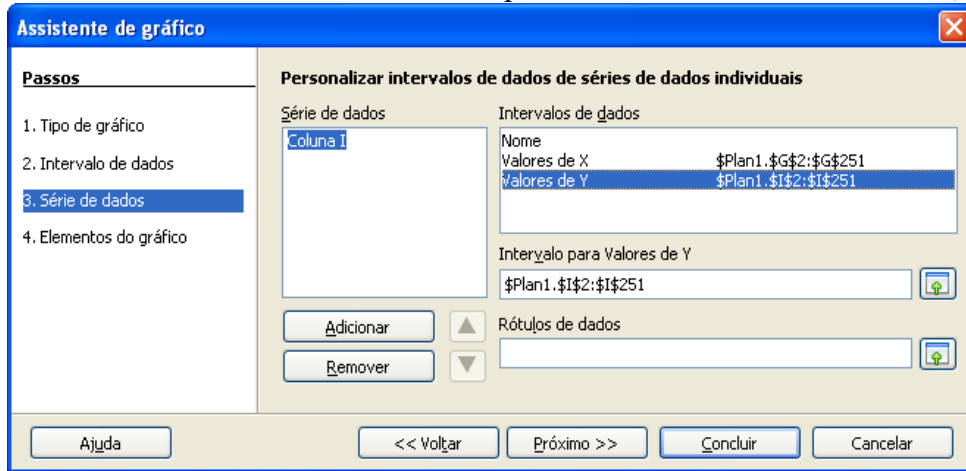


Figura 87 – Assistente gráfico do Calc – diagrama de dispersão – 3ª parte – dados prontos

Fonte: adaptada pelo autor de Sun Microsystems ®

Pressionando “Próximo” podemos finalizar o gráfico: inserir título, nomes dos eixos, e desmarcar a opção “Exibir legenda”, pois não há necessidade.

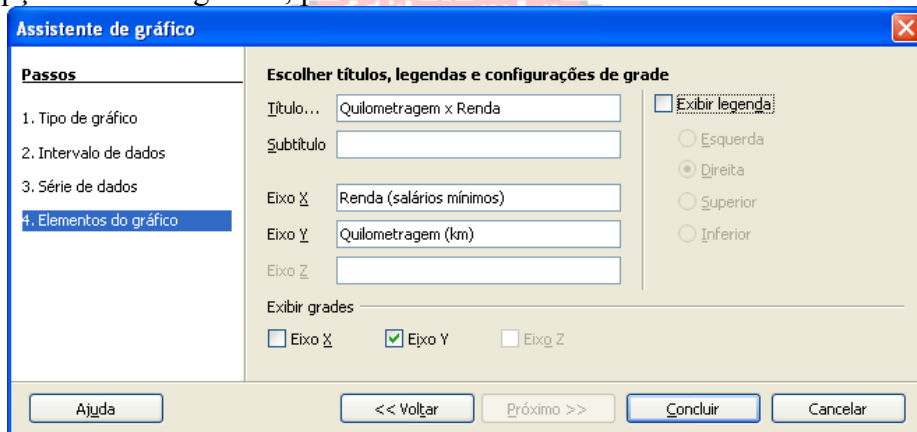


Figura 88 – Assistente gráfico do Calc – diagrama de dispersão – 4ª parte

Fonte: adaptada pelo autor de Sun Microsystems ®

Ao pressionar “Concluir” surgirá o gráfico mostrado na figura 84.

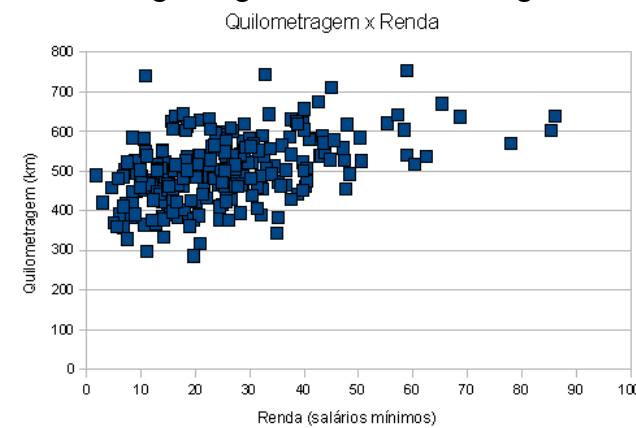


Figura 89 – Quilometragem por Renda

Fonte: adaptada pelo autor de Sun Microsystems ®

Temos que modificar a escala do eixo Y. Para modificar o gráfico selecione-o e pressione o botão direito do mouse, levando à figura 85 (veja figuras 18 e 48 também):

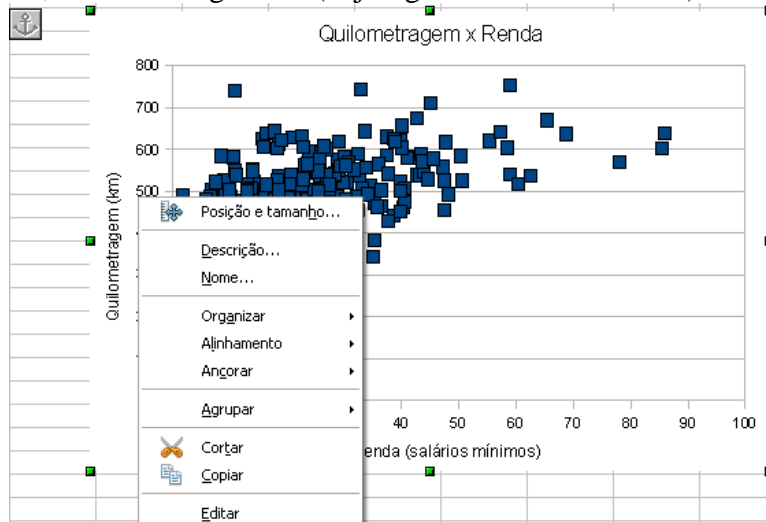


Figura 90 – Opções do gráfico para o diagrama de dispersão Quilometragem por Renda

Fonte: adaptada pelo autor de Sun Microsystems ®

Escolhemos “Editar”. Para mudar a escala do eixo Y é preciso colocar o cursor exatamente sobre o eixo Y, e pressionar o botão esquerdo do mouse levando à figura 86.

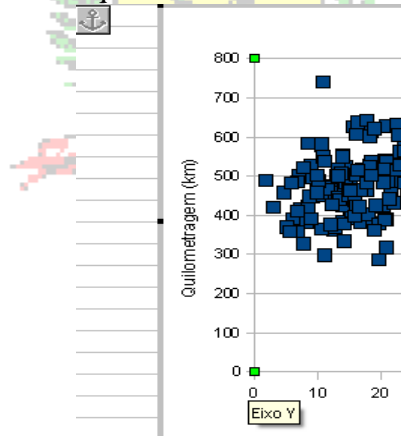


Figura 91 - Seleção do eixo Y

Fonte: adaptada pelo autor de Sun Microsystems ®

Posteriormente, mantendo o cursor sobre o eixo e pressionando o botão direito do mouse vamos ter acesso às opções relativas ao eixo Y, como mostrado na figura 87.

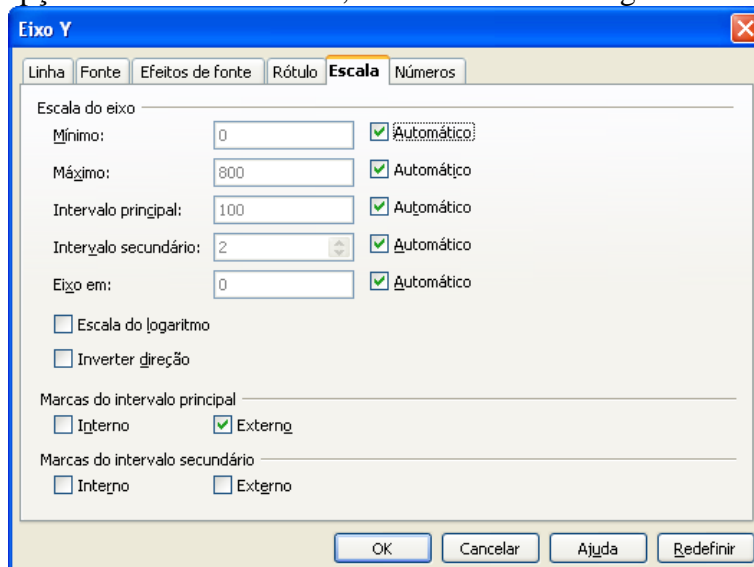


Figura 92 - Opções de formatação de eixo: escala

Fonte: adaptada pelo autor de Sun Microsystems ®

O comportamento padrão do Calc é construir a escala do gráfico com os valores mínimo e máximo encontrados nos dados. Mas algumas vezes, como no nosso problema, isso pode ser modificado, levando a um gráfico em que os dados estão muito concentrados. Como TODOS os valores de quilometragem estão acima de 200 km, vamos mudar o "Mínimo" da escala para 200, o que pode ser visto na figura 88.

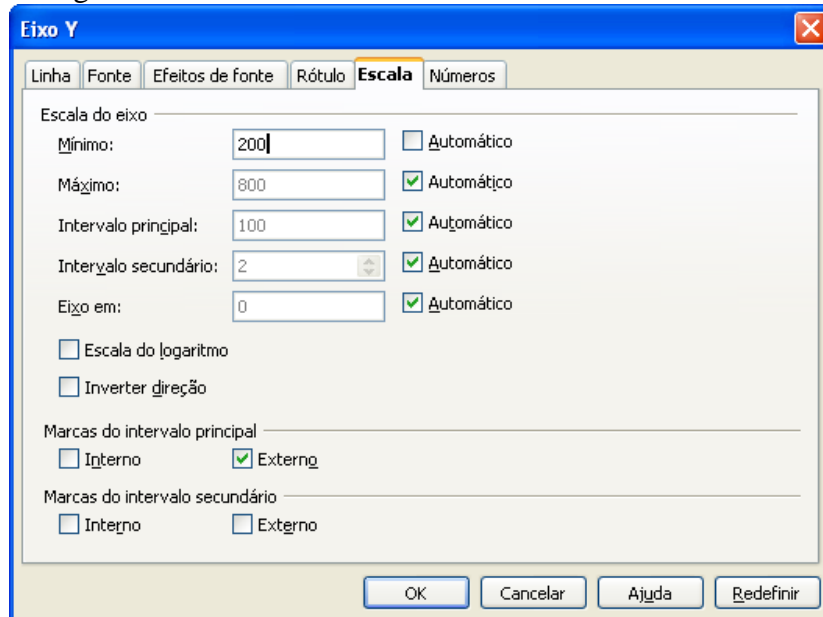


Figura 93 - Formatação de eixo: escala modificada
Fonte: adaptada pelo autor de Sun Microsystems ®

Pressionando "OK" vamos chegar ao gráfico mostrado na figura 89.

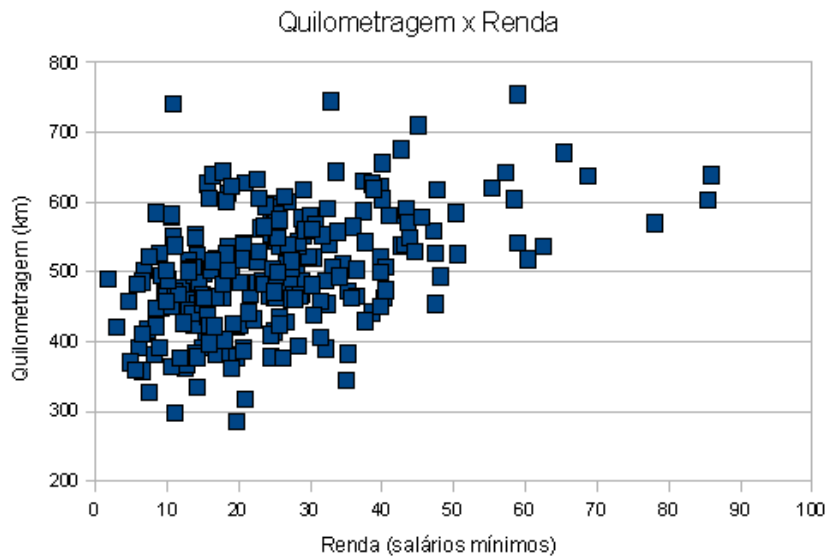


Figura 94 - Diagrama de dispersão Quilometragem por Renda
Fonte: adaptada pelo autor de Microsoft ®

Agora podemos fazer uma análise do diagrama de dispersão:

- a força da correlação não parece ser muito grande, pois os pontos estão relativamente distantes entre si.
 - a correlação entre elas parece ser positiva, pois se observa que a nuvem de pontos tem um comportamento crescente, ou seja, maiores valores de renda, maiores valores de quilometragem (o que é plausível).
 - quanto à forma do relacionamento, isto é, que tipo de curva poderíamos ajustar aos dados para realização de previsões, não parece ser algo muito fácil decidir. Poderíamos ajustar vários modelos aos dados, mas a aderência fosse ruim devido à fraca correlação.
- “Existe correlação entre renda e quilometragem”? Sim, mas ela não é forte, e aparenta ser positiva.