

Procedimentos de Análise Exploratória de Dados utilizando o Excel 2003

Professor Marcelo Menezes Reis

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 Excel, versão 2003. 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. Figuras adaptadas pelo autor de Microsoft ®, exceto onde indicado.

A Megamontadora TOYORD LINK (trata-se de uma empresa fictícia, e de uma pesquisa fictícia) 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 maio de 2015: 250 clientes foram entrevistados entre o total de 30.000 que compraram veículos novos entre maio de 2014 e maio de 2015. A pesquisa foi restringida aos modelos mais vendidos, e que já estão no mercado há 10 anos. As seguintes variáveis foram obtidas:

- Modelo comprado: o compacto Chiconaulta, o sedã médio DeltaForce3, a perua familiar Valentiniana, a van SpaceShuttle ou o luxuoso LuxuriousCar.
- 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.
- Opinião sobre o design: se os clientes consideram o design do veículo comprado ultrapassado, atualizado, ou adiante dos concorrentes.
- 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.
- Opinião geral sobre o veículo adquirido: muito insatisfeito, insatisfeito, satisfeito, bastante satisfeito.
- Renda declarada pelo cliente: em salários mínimos mensais.
- Número de pessoas geralmente transportadas no veículo.
- Quilometragem mensal média percorrida com o veículo.
- 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).
- Idade do cliente em anos completos.

Conteúdo

1. Procedimentos para variáveis qualitativas	4
1.1 – Tabela de frequências e gráfico da variável Modelo (uma variável qualitativa)	4
1.2 – Tabela de frequências e gráfico do cruzamento das variáveis Modelo e Opinião Geral sobre o veículo (duas variáveis qualitativas)	17
2. Procedimentos para variáveis quantitativas	22
2.1 – Procedimentos para variáveis quantitativas discretas	22
2.2 – Procedimentos para variáveis quantitativas contínuas	27
2.2.1 – Tabela de frequências agrupada em classes	27
2.2.2 – Categorização de uma variável quantitativa contínua	32
2.3 – Cálculo de Medidas de Síntese	35
2.3.1 – Cálculo de Medidas de Síntese através de inserção de funções	35
2.3.2 – Cálculo de Medidas de Síntese utilizando tabelas dinâmicas	36
2.3 – Procedimentos para duas variáveis quantitativas (diagrama de dispersão)	38

Índice de Figuras

Figura 1 - Opção de tabela dinâmica	4
Figura 2 - Etapa 1 da tabela dinâmica	4
Figura 3 - Etapa 2 da tabela dinâmica	5
Figura 4 - Tela de seleção de intervalo de dados – minimizada	5
Figura 5 - Etapa 3 da tabela dinâmica	5
Figura 6 - Layout da tabela dinâmica	6
Figura 7 - Tabela dinâmica para a variável Modelo	6
Figura 8 - Ações possíveis na tabela dinâmica	7
Figura 9 - Tabela dinâmica para a variável Modelo	7
Figura 10 - Ações possíveis na tabela dinâmica	7
Figura 11 - Mudança na Opção "Mostrar Dados"	8
Figura 12 - Opções da etapa final da tabela dinâmica	8
Figura 13 - Etapa final da tabela dinâmica	8
Figura 14 - Tabela dinâmica da variável Modelo	9
Figura 15 - Opções de exibição para Modelo	9
Figura 16 - Tabela dinâmica com frequências e percentuais para a variável Modelo	9
Figura 17 - Ações possíveis para a tabela dinâmica	10
Figura 18 - Gráfico dinâmico para a variável Modelo - 1a etapa	10
Figura 19 - Opções de dados no gráfico dinâmico da variável Modelo	11
Figura 20 - Gráfico dinâmico para a variável Modelo - 2a etapa	11
Figura 21 - Opções de gráfico dinâmico	11
Figura 22 - Gráfico dinâmico para a variável Modelo - 3a etapa	12
Figura 23 - Seleção do título do gráfico	12
Figura 24 - Formatação do título do gráfico	12
Figura 25 - Opções para formatar o título do gráfico	13
Figura 26 - Opções de modificação da fonte do título do gráfico	13
Figura 27 - Seleção do eixo vertical	13
Figura 28 - Opções de formatação de eixo	14
Figura 29 - Gráfico em barras para a variável Modelo	14
Figura 30 - Opções de formatação do gráfico	15
Figura 31 - Tipos de gráfico	15
Figura 32 - Opções de gráfico em setores ("pizza").	15
Figura 33 - Gráfico em setores para a variável Modelo	16
Figura 34 - Opções de gráfico em setores	16
Figura 35 - Opções de legenda em gráfico de setores	16
Figura 36 - Opções de rótulos de dados no gráfico em setores	17
Figura 37 - Gráfico em setores para a variável Modelo	17
Figura 38 - Aviso sobre aproveitamento de memória	18
Figura 39 - Opção de qual tabela dinâmica prévia usar	18
Figura 40 - Layout da tabela dinâmica de Modelo x Opinião Geral	18
Figura 41 - Mudança na Opção "Mostrar Dados"	19
Figura 42 - Tabela de contingências: Modelo por Opinião geral sobre o veículo	19

Figura 43 - Gráfico em barras múltiplas da Opinião Geral por Modelo	20
Figura 44 - Opções de gráfico dinâmico	21
Figura 45 - Opções de gráfico dinâmico - colunas 100% empilhadas	21
Figura 46 - Gráfico de colunas 100% empilhadas: Opinião geral por Modelo	21
Figura 47 - Tabela de frequência: anos de Remodelação	22
Figura 48 - Tabela de frequência: Anos de educação	22
Figura 49 - Assistente gráfico - etapa 1	23
Figura 50 - Assistente gráfico: etapa dos dados de origem	23
Figura 51 - Assistente gráfico: etapa 3.	24
Figura 52 - Assistente gráfico: etapa 4.	24
Figura 53 - Gráfico de anos de remodelação - 1a versão	24
Figura 54 - Gráfico de anos de remodelação - 2a versão	25
Figura 55 - Opções de gráficos	25
Figura 56 - Opções de formatação da série de dados	26
Figura 57 - Opções para as colunas	26
Figura 58 - Histograma de anos de remodelação	27
Figura 59 - Barra de ferramentas do Excel: "Colar Função"	27
Figura 60 - Funções disponíveis no Excel	28
Figura 61 - Funções estatísticas	28
Figura 62 - Função Mínimo	29
Figura 63 - Valores mínimo e máximo de Renda	29
Figura 64 - Amplitude dos valores de Renda	29
Figura 65 - Cálculo dos limites inferior e superior da primeira classe de renda	30
Figura 66 - Cálculo dos limites inferior e superior da segunda classe de renda	30
Figura 67 - Classes de renda	30
Figura 68 - Obtenção da frequência da primeira classe de renda	31
Figura 69 - Obtenção da frequência da última classe de renda	31
Figura 70 - Obtenção dos pontos médios das classes de renda	31
Figura 71 - Histograma das Rendas anuais (agrupados em classes)	32
Figura 72 - Função lógica SE	32
Figura 73 - Argumentos da função SE	33
Figura 74 - Procedimento para categorizar a variável Renda	33
Figura 75 - Categorização da variável Renda no Excel ®	34
Figura 76 - Variável Renda Categorizada	34
Figura 77 - Tabela dinâmica da variável Renda Categorizada	34
Figura 78 - Medidas de síntese da variável Renda	35
Figura 79 - Inserção de função: Média	36
Figura 80 - Tabela dinâmica: layout de medidas de síntese de Renda por Modelo	36
Figura 81 - Renda em função do Modelo	37
Figura 82 - Assistente gráfico - 1a etapa	38
Figura 83 - Assistente gráfico - 2a etapa	39
Figura 84 - Assistente gráfico: adição de sequências - início	39
Figura 85 - Assistente gráfico: adição de sequências - completo	39
Figura 86 - Assistente gráfico - 3a etapa	40
Figura 87 - Diagrama de dispersão de quilometragem por renda	40
Figura 88 - Seleção do eixo Y	41
Figura 89 - Opções de formatação de eixo: escala Figura 90 - Formatação de eixo: escala modificada	41
Figura 91 - Diagrama de dispersão de Quilometragem por Renda	41

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 Excel 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 Excel. Procure pela opção "Relatório de tabela dinâmica", veja a Figura 1. Escolha esta opção.

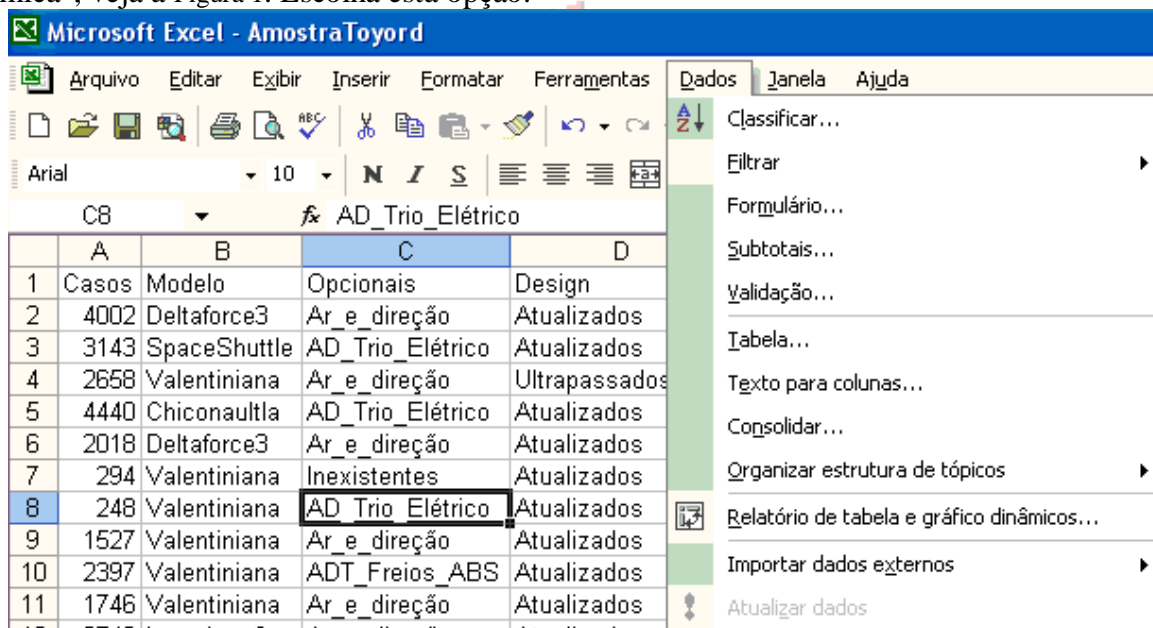


Figura 1 - Opção de tabela dinâmica

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

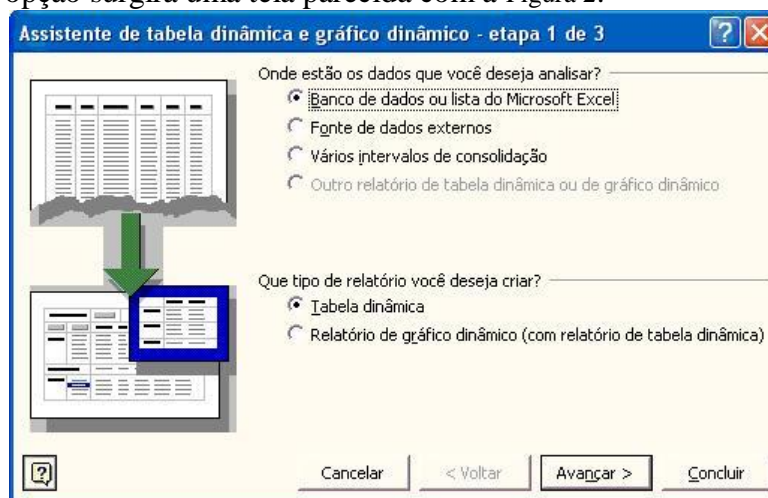


Figura 2 - Etapa 1 da tabela dinâmica

Trata-se da etapa 1 de 3. Você precisa informar onde estão os dados de seu interesse. Usualmente estão na própria planilha do Excel, podem, porém ser apenas uma variável ou conjunto de variáveis.

Neste exemplo, os dados realmente estão em uma lista do Microsoft Excel, mantenha como está e pressione "Avançar", o que levará à Figura 3.

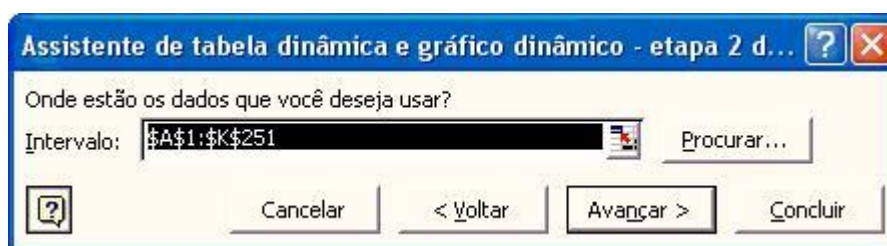


Figura 3 - Etapa 2 da tabela dinâmica

c) Neste momento é preciso especificar exatamente onde estão os dados da tabela dinâmica: um intervalo de células do Excel, indicando as colunas (marcadas com letras) e as linhas (marcadas com números). Muitas vezes, como na Figura 3, o Excel automaticamente seleciona todos os dados disponíveis na planilha (no caso da célula A1 até a célula K251, compreendendo as 250 observações de cada uma das 11 variáveis). Contudo, se isso não ocorrer, você pode selecionar o intervalo por conta própria, de duas maneiras:

- digitando no campo "Intervalo" (ver Figura 3) as referências das células (A1:K251, o : significa que todas as células do intervalo serão consideradas).
- buscando e selecionando o intervalo na planilha; para tanto você precisa pressionar a pequena seta vermelha imediatamente à direita do intervalo, o que resultará na Figura 4:

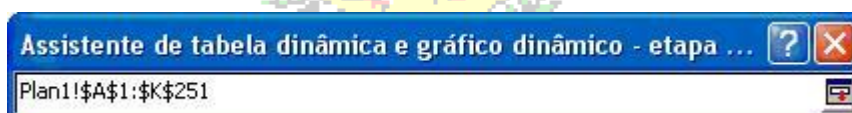


Figura 4 - Tela de seleção de intervalo de dados – minimizada

Com a seleção neste formato você pode procurar o intervalo desejado e selecioná-lo utilizando o mouse. Para retornar à situação da Figura 3 basta pressionar a pequena seta vermelha novamente. Pressionando "Avançar" na tela da Figura 3 chegamos à Figura 5.

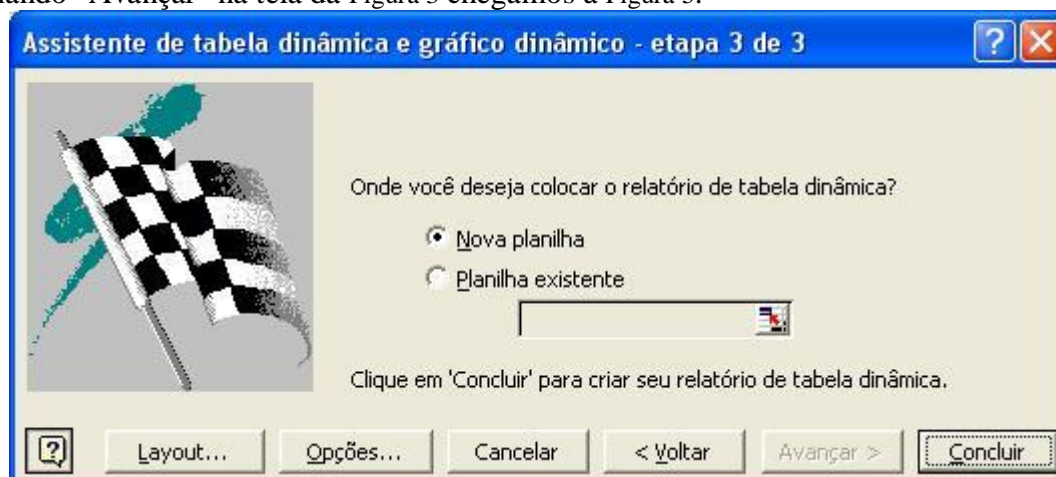


Figura 5 - Etapa 3 da tabela dinâmica

d) Uma vez selecionados os dados é preciso definir o layout da tabela dinâmica, como os dados estarão distribuídos, e que ação a tabela dinâmica deverá executar. Na Figura 5 você deve ter percebido o botão "Layout". Pressione-se para definir o layout, e chegaremos à Figura 6.

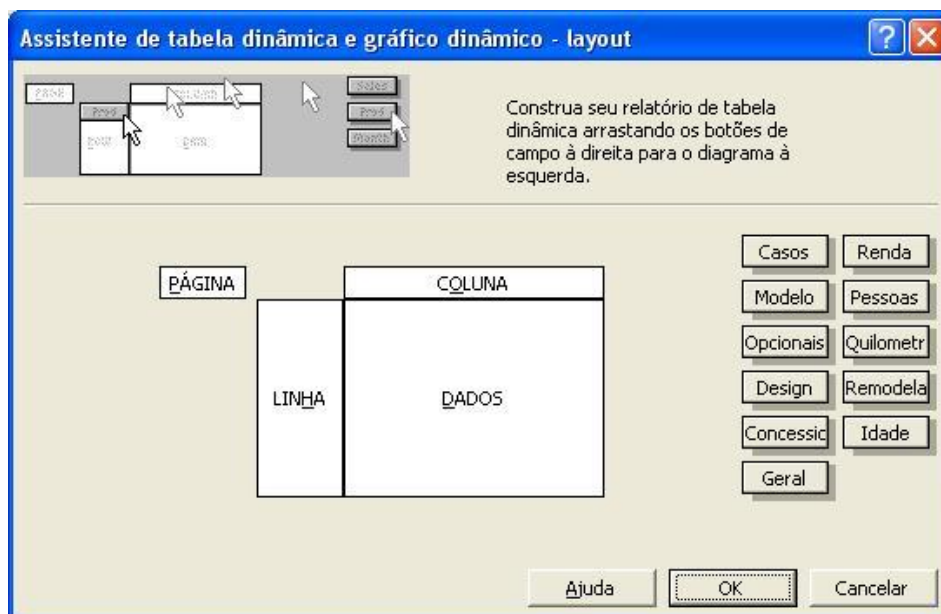


Figura 6 - Layout da tabela dinâmica

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 para variáveis qualitativas é a contagem dos valores, tal como mostrado na Figura 7. Já para variáveis quantitativas, a ação padrão é a soma.

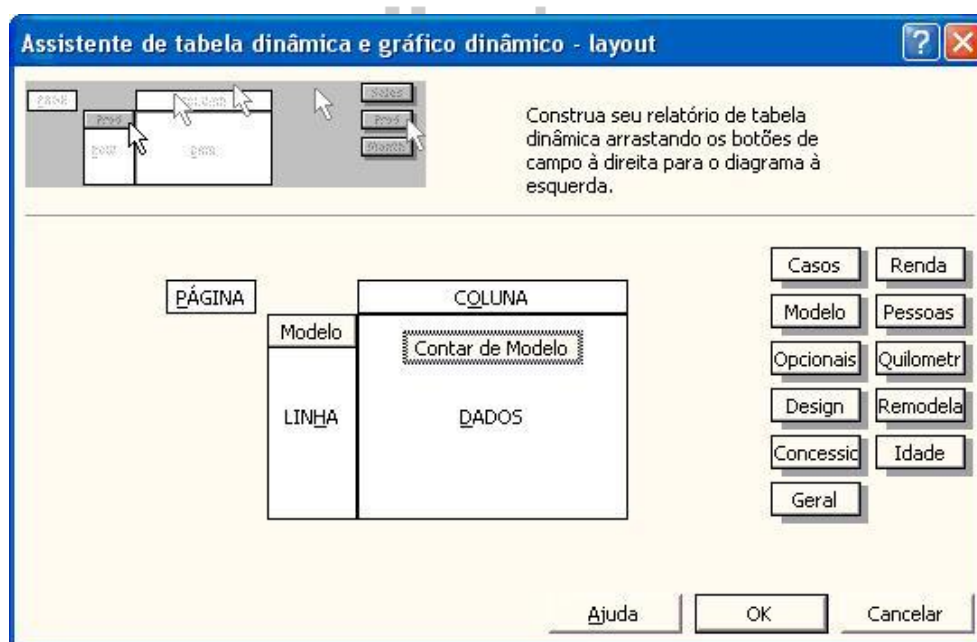


Figura 7 - Tabela dinâmica para a variável Modelo

Se você pressionar duas vezes "Contar de Modelo" surgirá a tela exposta na Figura 8, onde estão as várias ações possíveis (como média, mínimo, máximo, desvio padrão, que serão úteis quando formos analisar variáveis quantitativas):

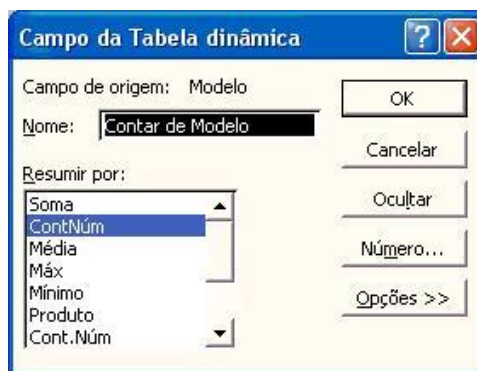


Figura 8 - Ações possíveis na tabela dinâmica

No presente caso queremos realmente contar o número de ocorrências dos valores da variável Modelo, portanto nada é mudado, e pressionamos "OK" (na tela da Figura 8). Vamos supor também que queremos que a tabela contenha também os valores dos percentuais de cada modelo. Para tanto, é preciso selecionar mais uma vez a variável Modelo (na Figura 7) e arrastá-la até o campo Dados, o que resultará na Figura 9:

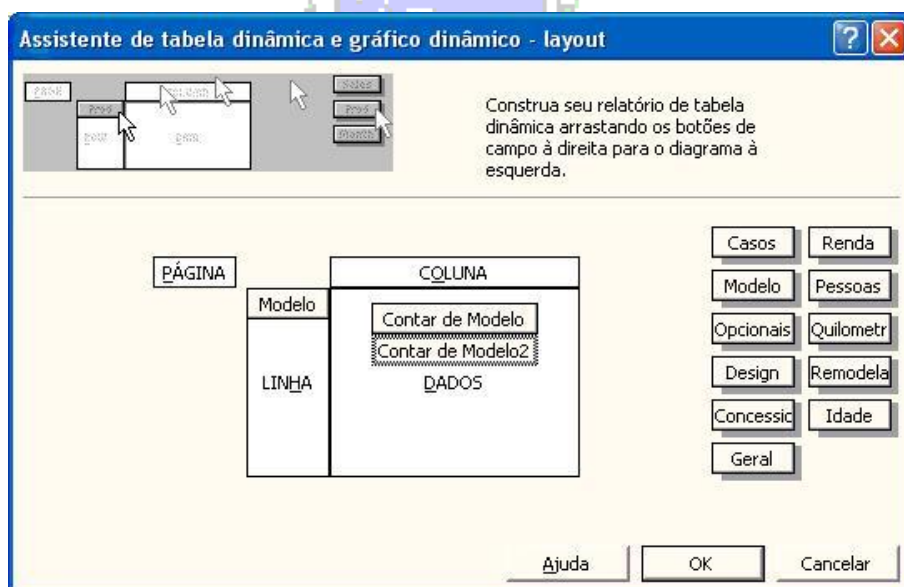


Figura 9 - Tabela dinâmica para a variável Modelo

O botão Contar de Modelo2 precisa ser alterado para percentual. Clicando duas vezes sobre ele teremos:



Figura 10 - Ações possíveis na tabela dinâmica

Precisamos pressionar o botão "Opções", para mudar a forma como os valores são registrados, de contagem simples para contagem percentual. O que resultará na sequência de operações a seguir.



Figura 11 - Mudança na Opção "Mostrar Dados"

A opção padrão é Normal, que significa que os dados serão mostrados como contagem de frequência (tela à esquerda na Figura 11). Se selecionarmos o botão ao lado de Normal surge a tela central da Figura 11: observe que há várias maneiras de apresentar os dados, incluindo percentuais: como precisamos mostrar os percentuais em relação ao total da variável Modelo escolhemos a opção "% do total" (tela à direita na Figura 11).

Pode-se então pressionar "OK" na tela à direita da Figura 11, e depois "OK" novamente (na tela da Figura 9), levando à mesma situação da Figura 5.

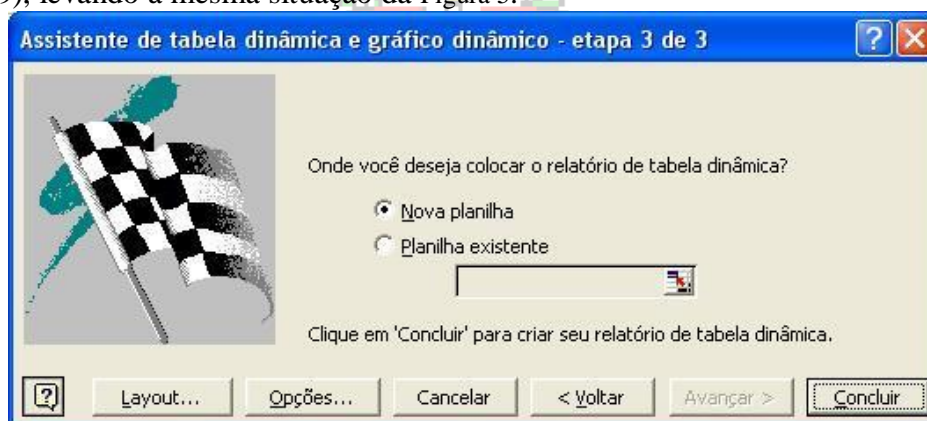


Figura 12 - Opções da etapa final da tabela dinâmica

e) É preciso definir onde você quer os resultados: em uma nova planilha (padrão), ou na planilha existente, precisando especificar a célula a partir da qual os resultados devem ser postos. No nosso caso, podemos pôr os resultados a partir da célula N2, tal como mostrado na Figura 13.

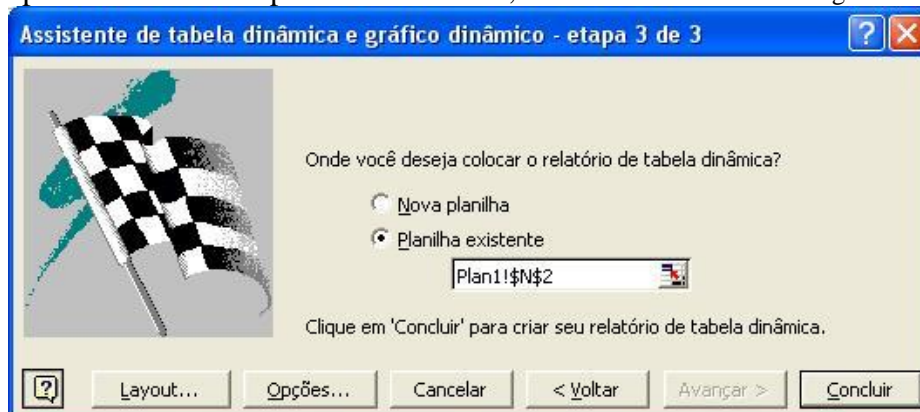


Figura 13 - Etapa final da tabela dinâmica

Após definir onde colocar a tabela dinâmica, e feitas ou não as modificações na sua apresentação, basta pressionar "Concluir", o resultado deverá ser:

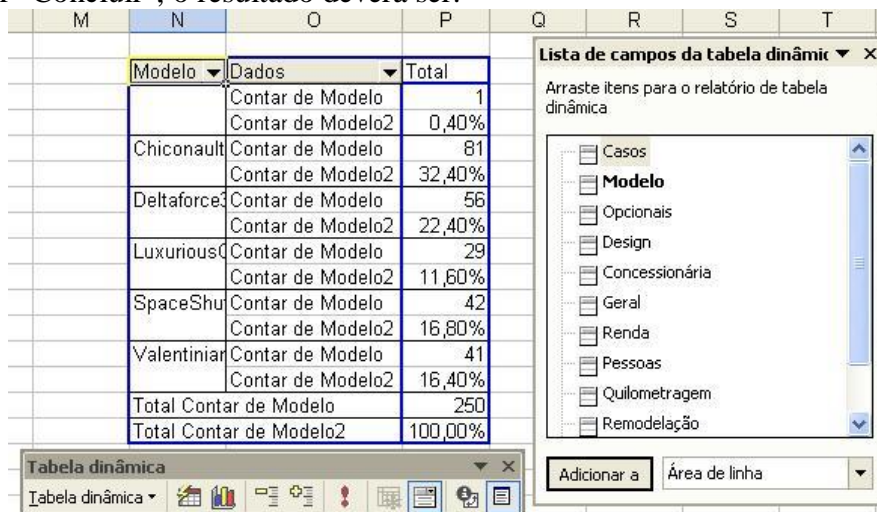


Figura 14 - Tabela dinâmica da variável Modelo

Surgem duas barras de ferramentas: Tabela Dinâmica e Lista de campos da tabela dinâmica. Podemos clicar sobre X em ambos para desativá-los. Observe que a tabela contém as frequências e percentuais 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. Obviamente isso não nos interessa, então podemos removê-lo da tabela: clique sobre o campo Modelo na tabela, e surgirá a seguinte tela:

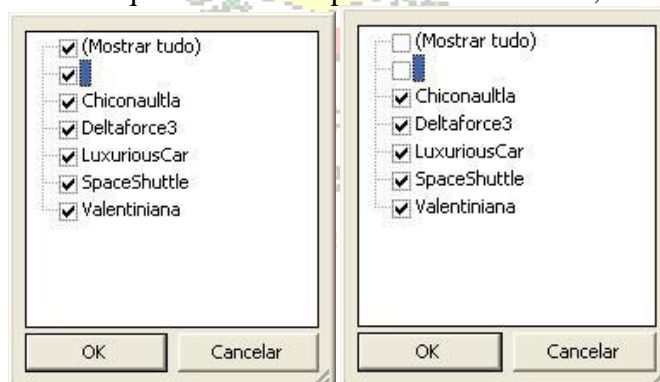



Figura 15 - Opções de exibição para Modelo

O padrão de exibição é mostrar todos os valores da variável, como está na tela à esquerda da Figura 15, incluindo erros e células vazias. Como queremos eliminar as células vazias basta desmarcar a opção correspondente acima, resultando na tela à direita da Figura 15. Pressionando "OK" a tabela passará a ser:

N	O	P
Modelo ▼	Dados ▼	Total
Chiconault	Contar de Modelo	81
	Contar de Modelo2	32,53%
Deltaforce3	Contar de Modelo	56
	Contar de Modelo2	22,49%
Luxurious0	Contar de Modelo	29
	Contar de Modelo2	11,65%
SpaceShur	Contar de Modelo	42
	Contar de Modelo2	16,87%
Valentiniar	Contar de Modelo	41
	Contar de Modelo2	16,47%
Total Contar de Modelo		249
Total Contar de Modelo2		100,00%

Figura 16 - Tabela dinâmica com frequências e percentuais para a variável Modelo

Com a tabela pronta é possível avaliar os resultados. É possível também construir gráficos com base na tabela. Há duas formas de fazer isso:

- pressionando o ícone do "Assistente Gráfico": .

- ou, já que estamos lidando com uma tabela dinâmica, selecionar a tabela e pressionar o botão direito do mouse, o que nos levará à Figura 17.

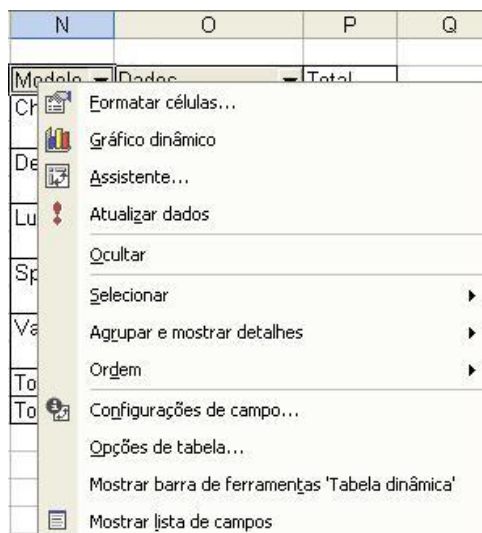


Figura 17 - Ações possíveis para a tabela dinâmica

Dentre as várias opções podemos escolher "Gráfico dinâmico", e o Excel automaticamente construirá o gráfico padrão, em colunas, tal como o da Figura 18, criando automaticamente uma nova planilha no arquivo.

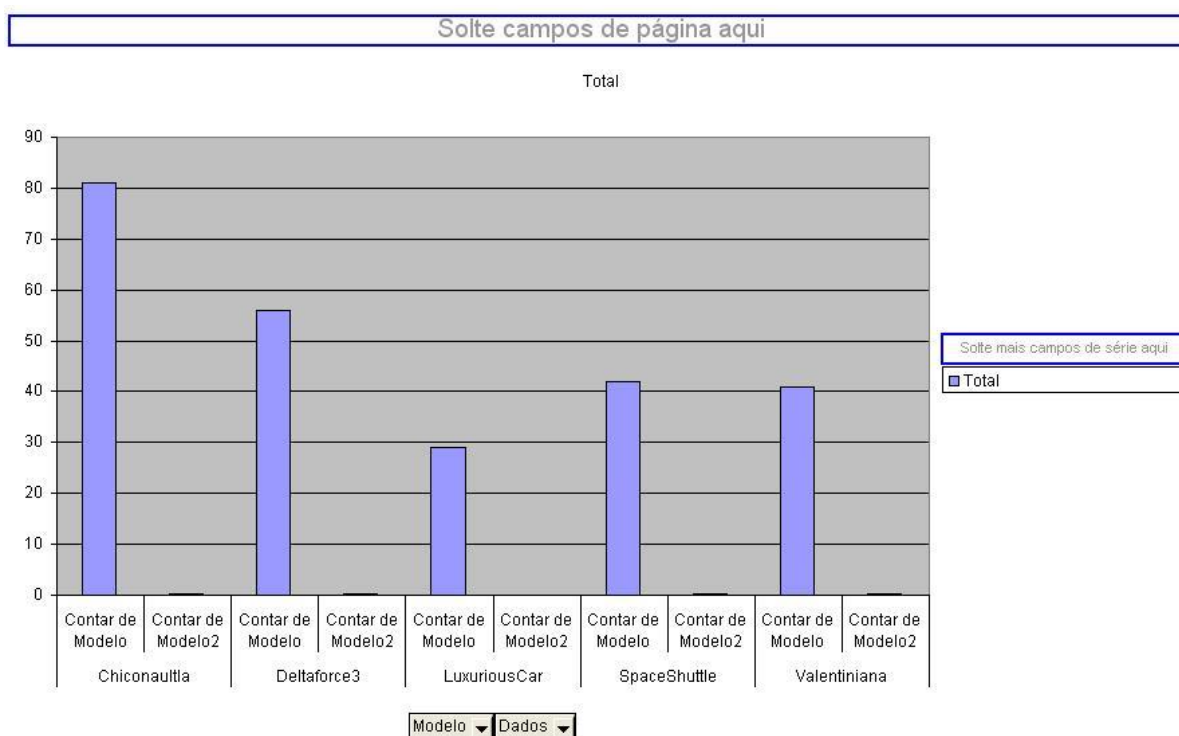


Figura 18 - Gráfico dinâmico para a variável Modelo - 1a etapa

Percebemos claramente que o gráfico está um pouco "confuso": há expressões do tipo "Solte mais campos...", e os dois registros da variável modelo (frequências e percentuais) estão no gráfico. Para tornar o gráfico mais limpo devemos tomar alguns cuidados:

- primeiramente, eliminar uma das contagens, no nosso caso vamos eliminar os percentuais; basta clicar sobre a opção Dados na Figura 18, e teremos a tela à esquerda da Figura 19; desmarcando a opção Contar de Modelo2, que registrava os percentuais, surgirá a tela à direita da Figura 19.

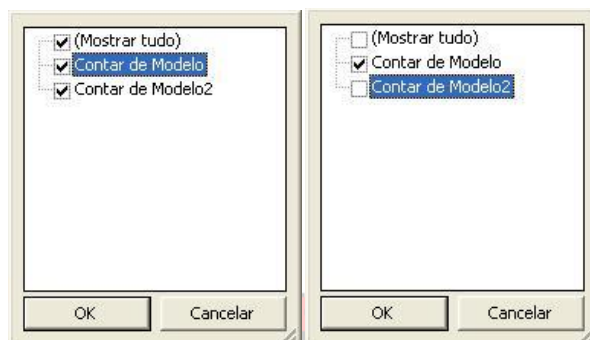


Figura 19 - Opções de dados no gráfico dinâmico da variável Modelo

O gráfico dinâmico passará a ser como o da Figura 20.

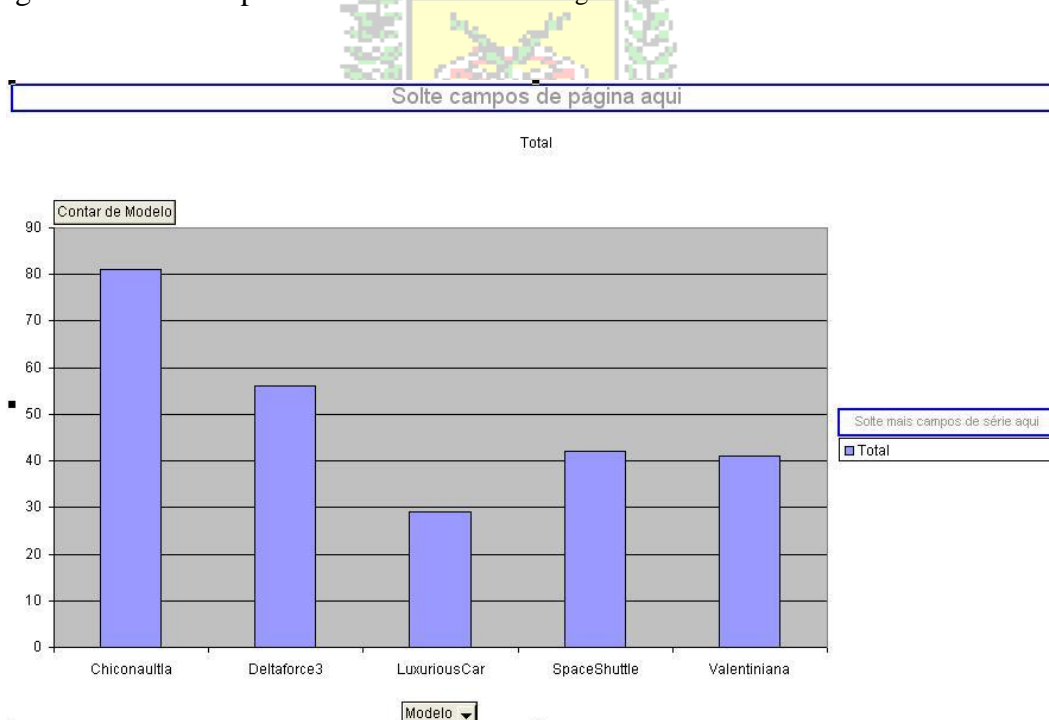


Figura 20 - Gráfico dinâmico para a variável Modelo - 2a etapa

- Em segundo lugar precisamos tirar as opções "Solte mais campos...". Basta selecionar o botão Contar de Modelo na Figura 20 e pressionar o botão direito do mouse, resultando na Figura 21.

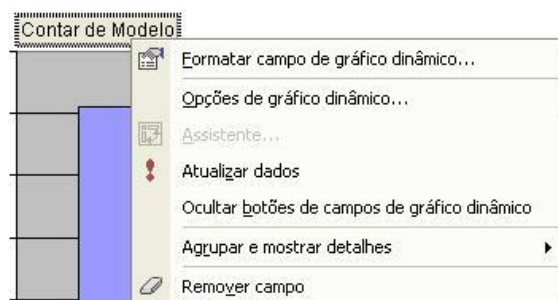


Figura 21 - Opções de gráfico dinâmico

Observe que há várias opções. Uma extremamente interessante é "Atualizar dados". Vamos supor que você tenha incluído ou removido dados, e modificado a tabela dinâmica que originou o gráfico: você pode atualizá-lo simplesmente selecionando esta opção. No nosso caso, queremos ocultar os botões de campos de gráfico dinâmico: basta selecionar esta opção, o que levará ao gráfico da Figura 22.

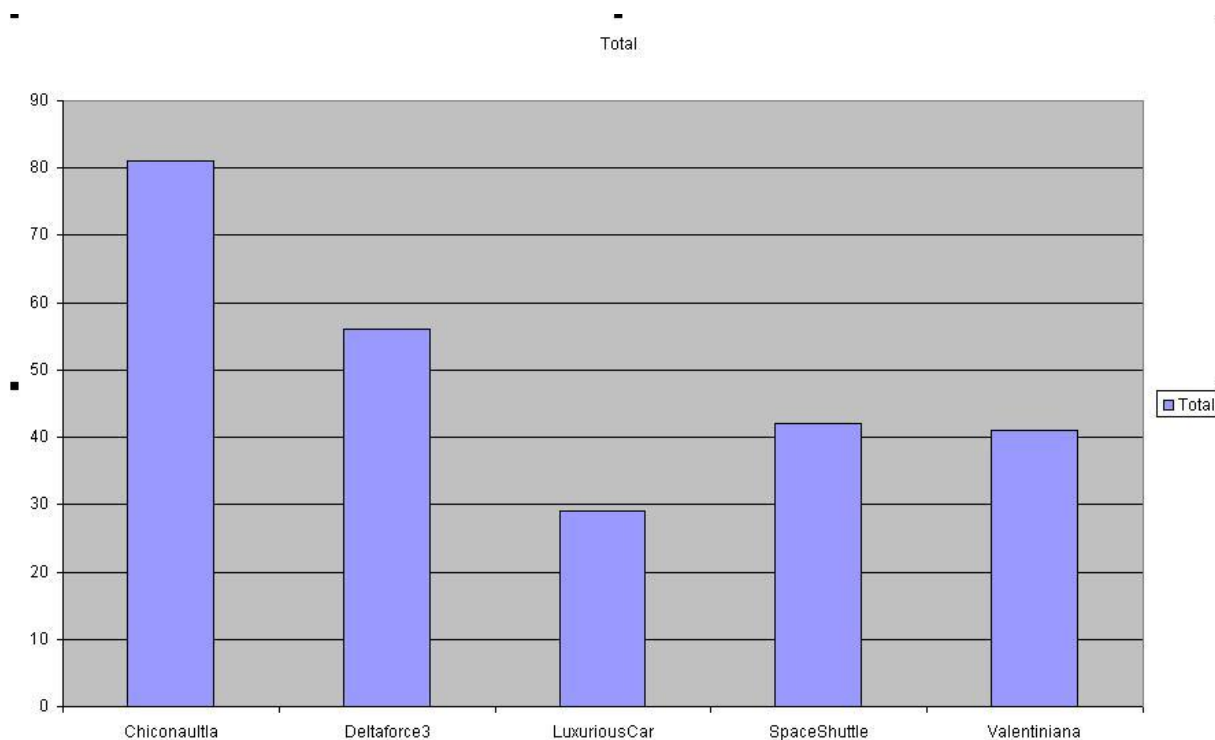


Figura 22 - Gráfico dinâmico para a variável Modelo - 3a etapa

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. Não há necessidade de legenda neste gráfico, pois há apenas uma variável.

Você pode modificar o título da seguinte maneira. Selecione o título diretamente, como na Figura 23.

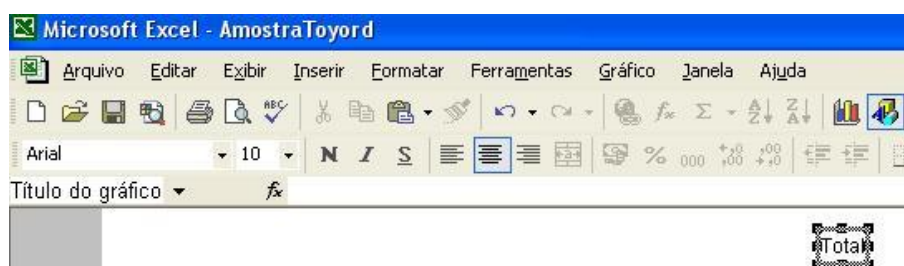


Figura 23 - Seleção do título do gráfico

Ao clicar duas vezes sobre o título você poderá modificá-lo diretamente. No nosso caso, como é um gráfico da variável Modelo, vamos trocar o título para "Modelos de veículos". Se quisermos mudar a fonte ou outro aspecto do título basta selecionar novamente o título e pressione o botão direito do mouse, resultando na Figura 24.



Figura 24 - Formatação do título do gráfico

Ao selecionar "Formatar título do gráfico" surgirá a tela da Figura 25.



Figura 25 - Opções para formatar o título do gráfico

Há várias opções disponíveis. Uma das mais usadas é a mudança da Fonte, para mudar tamanho, por negrito ou trocar o tipo da fonte. Escolhendo esta opção, teremos a tela da Figura 26:

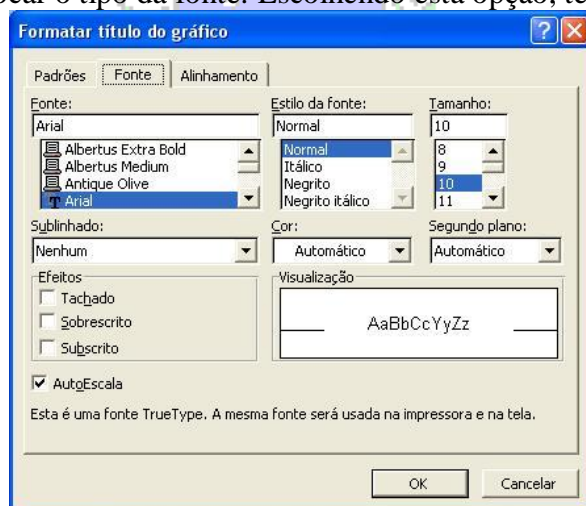


Figura 26 - Opções de modificação da fonte do título do gráfico

Por exemplo, poderíamos mudar "Estilo da fonte" para Negrito, e aumentar o tamanho da fonte para 14. Após fazer todas as modificações podemos pressionar "OK". Pressionando novamente "OK" na tela da Figura 25 terminamos as modificações no título.

Podemos modificar também os eixos do gráfico, o que pode ser muito útil quando quisermos mudar a sua escala para melhor visualização, ou para evitar os erros de interpretação mostrados na Unidade 3, Exemplo 2, Figuras 2 e 3. Basta selecionar o eixo, podemos escolher o eixo vertical do nosso gráfico, tal como mostrado na Figura 27.

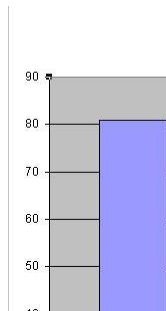


Figura 27 - Seleção do eixo vertical

Observe que o eixo está selecionado se aparecer o símbolo ■ no topo e na base do eixo. Pressionando o botão direito do mouse surgirá a opção “Formatar eixo”. Ao selecioná-la chegamos à tela da Figura 28.

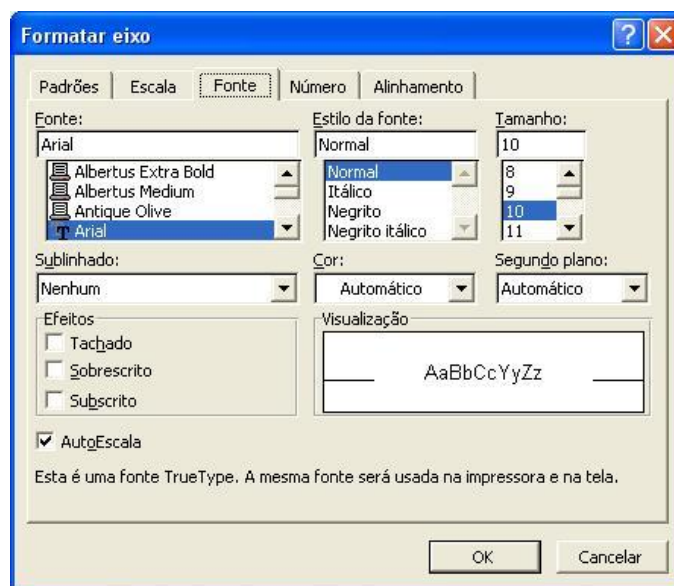


Figura 28 - Opções de formatação de eixo

Poderíamos mudar os padrões, a definição de número (reduzir ou aumentar o número de casas decimais), modificar a fonte, e mudar a escala. No nosso gráfico a escala já está correta, começando em zero, então não há necessidade de alteração. Mais à frente, quando estudarmos o caso de duas variáveis quantitativas, vamos mostrar como modificar a escala.

O gráfico de barras final está na Figura 29.

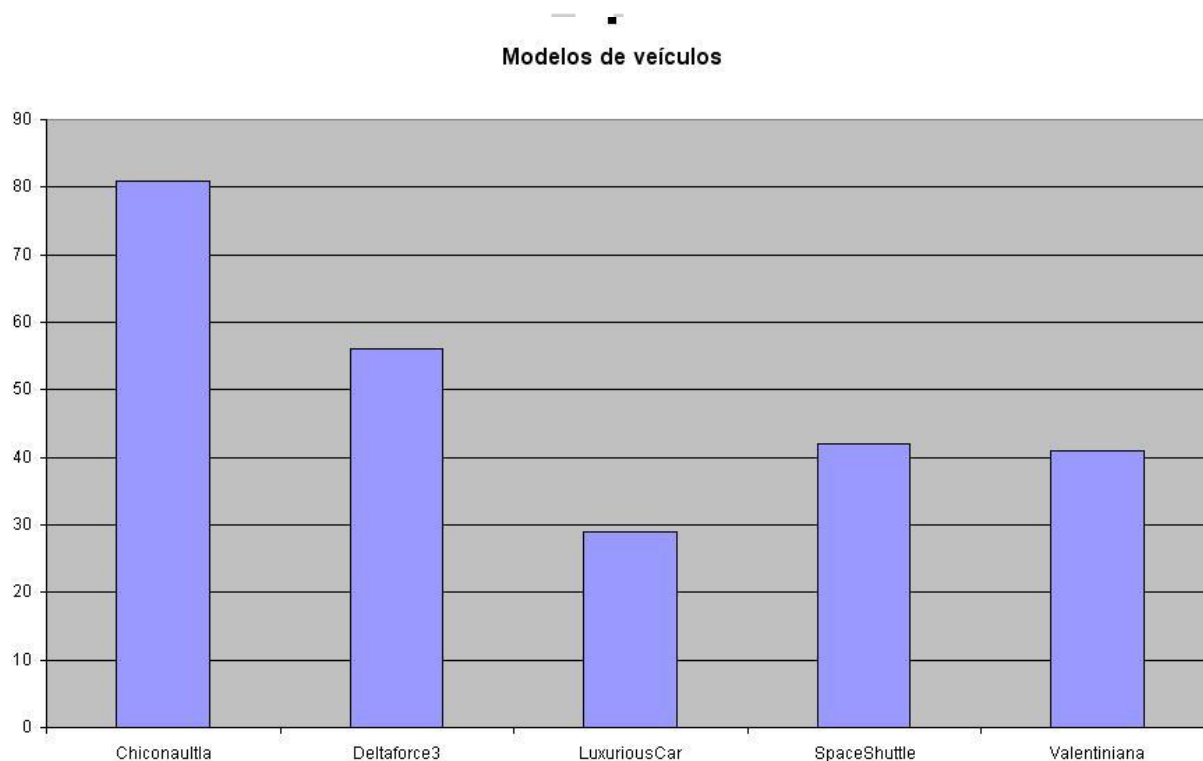


Figura 29 - Gráfico em barras para a variável Modelo

Com o gráfico pronto podemos fazer a interpretação: o modelo Chiconaultla é o mais vendido, seguido por DeltaForce3, SpaceShuttle, Valentiniana e LuxuriousCar. E se quiséssemos transformar o gráfico acima em um gráfico em setores? Não seria necessário refazer todo o processo. Basta selecionar o gráfico como um todo, pressionar o botão direito do mouse, e teremos a tela da Figura 30.

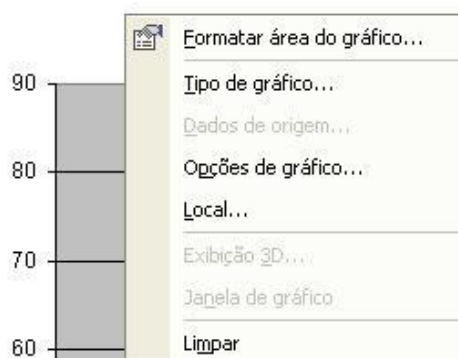


Figura 30 - Opções de formatação do gráfico

Podemos modificar vários aspectos do gráfico, entre eles o tipo de gráfico. Escolhendo esta opção chegamos à Figura 31.

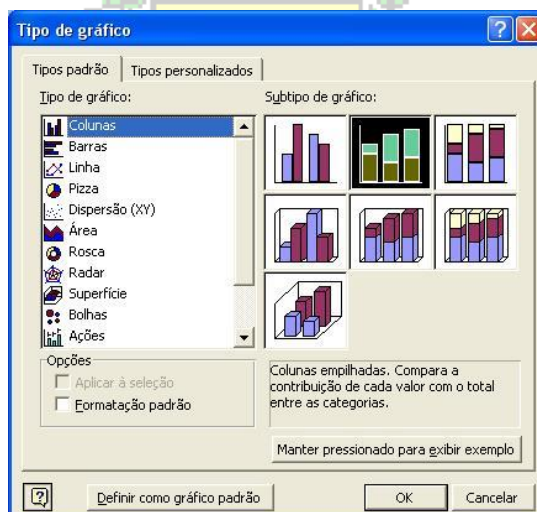


Figura 31 - Tipos de gráfico

O tipo padrão de gráfico é o de colunas (para nós barras), mas não queremos este. Podemos ver a opção "Pizza", com o formato do gráfico em setores que desejamos. Escolhendo esta opção vamos chegar à Figura 32.

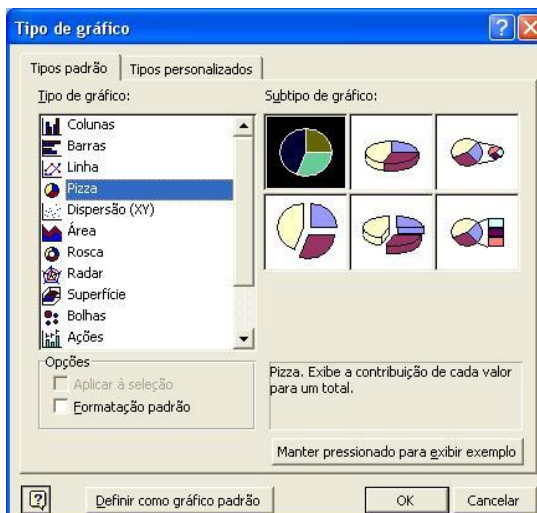


Figura 32 - Opções de gráfico em setores ("pizza").

Observe que há vários tipos de gráficos em setores, 2D e 3D. Vamos escolher a opção padrão. Automaticamente o Excel ® mudará o gráfico para o apresentado na Figura 33.

Modelos de veículos

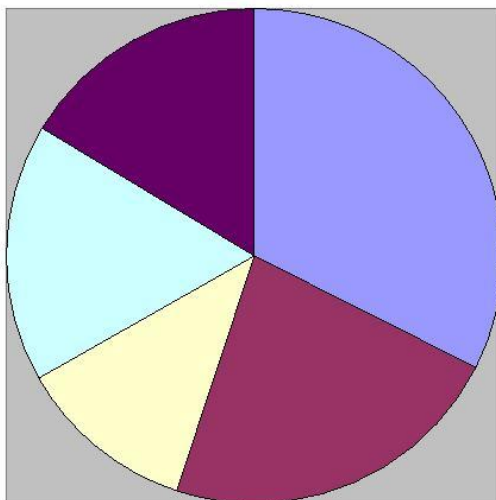


Figura 33 - Gráfico em setores para a variável Modelo

Observe que o gráfico praticamente não diz nada: não tem legenda e tampouco os percentuais de cada categoria. Precisamos acrescentá-los: selecione o gráfico e pressione o botão direito do mouse. Isso nos levará a uma tela semelhante a da Figura 30. Basta escolher “Opções de Gráfico”, e chegaremos à Figura 34.

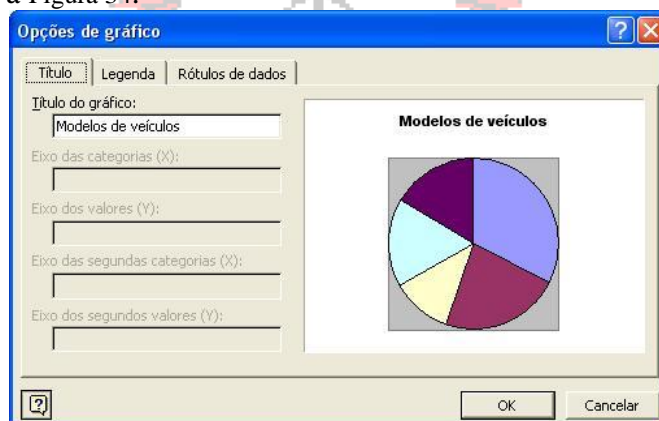


Figura 34 - Opções de gráfico em setores

O título pode ser modificado, e podemos modificar a legenda, escolhendo esta opção, tal como mostrado na Figura 35.



Figura 35 - Opções de legenda em gráfico de setores

Como queremos a legenda, marcamos a opção “Mostrar legenda”, e podemos escolher seu posicionamento, à direita, por exemplo. Mas, ainda faltam os percentuais associados a cada categoria, que precisamos marcar na opção “Rótulo de dados”, tal como na Figura 36.



Figura 36 - Opções de rótulos de dados no gráfico em setores

Há várias opções de rótulo. Como já inserimos a legenda podemos marcar apenas a “Porcentagem” associada a cada categoria do gráfico, levando ao resultado final (Figura 37)

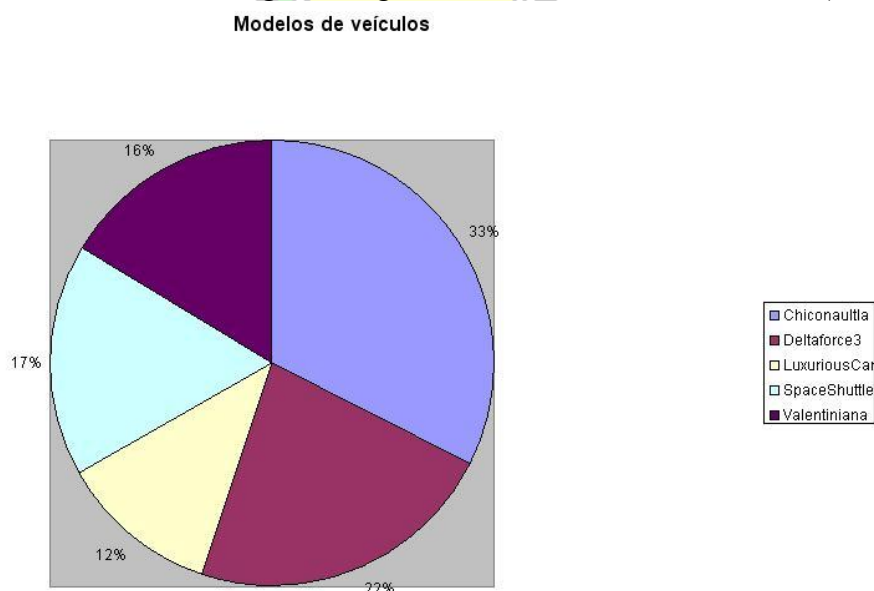


Figura 37 - Gráfico em setores para a variável Modelo

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%). Valentiniana 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 (Figura 1 a Figura 5) pode ser repetido literalmente. As diferenças começam a aparecer na determinação do layout da tabela. Se você estiver construindo a tabela após ter construído a tabela da variável Modelo, o Excel irá apresentar o seguinte aviso:

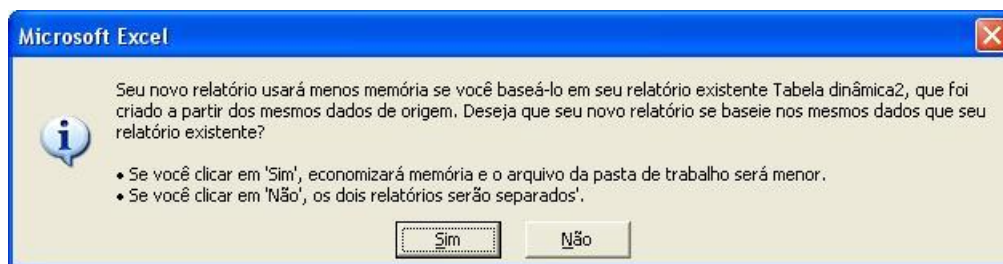


Figura 38 - Aviso sobre aproveitamento de memória

Como os dados de origem da tabela Modelo por Opinião Geral são os mesmos da tabela Modelo (porque nós escolhemos incorporar todos os dados da planilha), você pode optar por basear a nova tabela na existente, o que pode poupar alguma memória. Caso você escolha “Sim”, o Excel® apresentará todas as tabelas dinâmicas existentes, para que você possa selecionar qual é a de seu interesse (Figura 39).

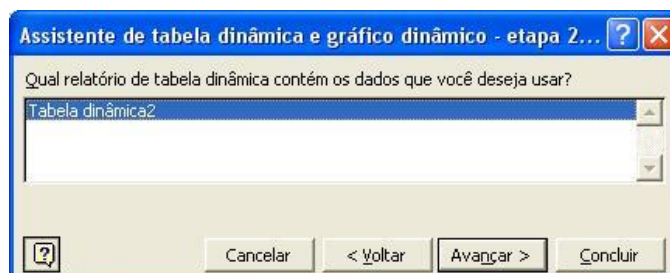


Figura 39 - Opção de qual tabela dinâmica prévia usar.

No nosso caso há apenas uma tabela dinâmica. Basta selecioná-la e pressionar “Avançar”, para que possamos definir o layout da nova tabela.

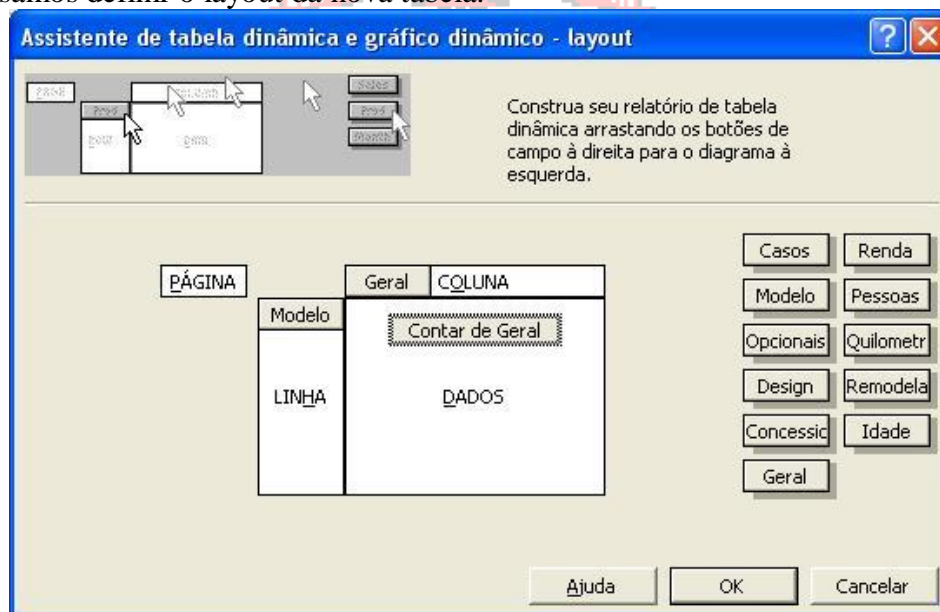


Figura 40 - Layout da tabela dinâmica de Modelo x Opinião Geral

Selecione e arraste a variável Modelo para o campo "LINHA", e a variável Geral (Opinião Geral) para o campo "COLUNA" (ou vice-versa). Selecione a variável Geral (ou a variável Modelo, mas APENAS uma delas) e arraste para o campo "DADOS": a ação padrão é "Contar de Geral" (ou "Contar de Modelo", se você escolher esta variável).

Podemos acrescentar percentuais à tabela também, mas aqui teremos três opções: percentuais calculados em relação aos totais das linhas, das colunas ou em relação ao total da tabela. Faremos isso de forma semelhante ao que fizemos no caso de apenas uma variável: “arrastando” a variável Geral para o campo Dados (ver Figura 9), três vezes, porque queremos obter três percentuais, e depois modificando as opções de “Mostrar dados”, como apresentado na Figura 41:

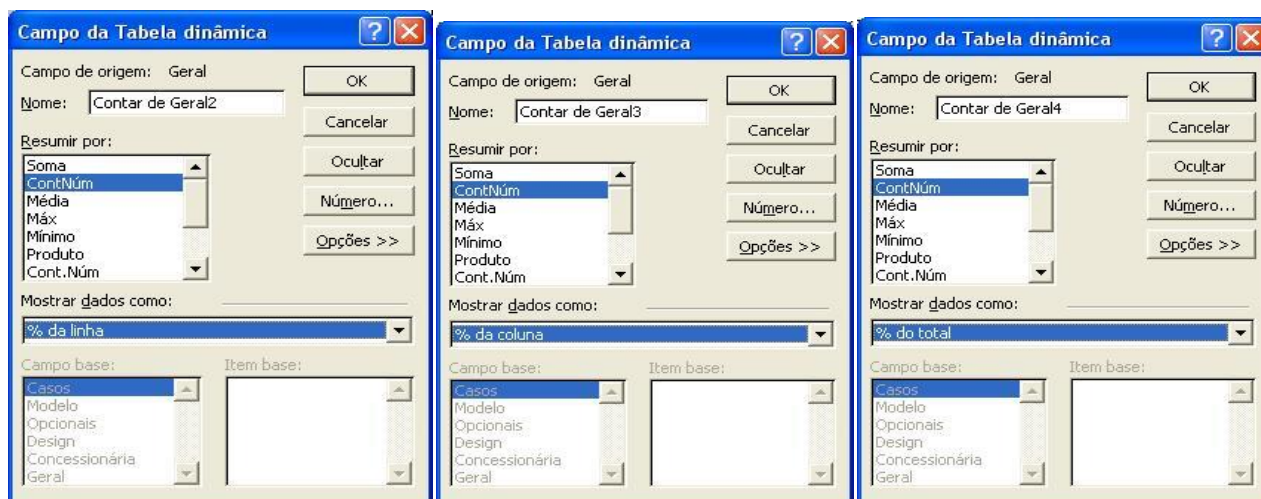


Figura 41 - Mudança na Opção "Mostrar Dados"

Na tela da esquerda da Figura 41 modificamos a exibição dos dados para a apresentação dos percentuais por linha, na tela central para percentuais por coluna, e na tela à direita para percentuais em relação ao total geral da tabela. Com estas modificações, pressionando “OK” nas telas da Figura 41, e no layout da tabela (Figura 40), e escolhendo onde os resultados da tabela serão colocados chegaremos à tabela final, mostrada na Figura 42:

		Geral				
Modelo	Dados	Bastante satisfeito	Insatisfeito	Muito insatisfeito	Satisfeito	Total geral
Chiconaultla	Contar de Geral		11	69	1	81
	Contar de Geral2	0,00%	13,58%	85,19%	1,23%	100,00%
	Contar de Geral3	0,00%	16,67%	62,73%	2,44%	32,53%
	Contar de Geral4	0,00%	4,42%	27,71%	0,40%	32,53%
Deltaforce3	Contar de Geral		22	29	5	56
	Contar de Geral2	0,00%	39,29%	51,79%	8,93%	100,00%
	Contar de Geral3	0,00%	33,33%	26,36%	12,20%	22,49%
	Contar de Geral4	0,00%	8,84%	11,65%	2,01%	22,49%
LuxuriousCar	Contar de Geral	19	1		9	29
	Contar de Geral2	65,52%	3,45%	0,00%	31,03%	100,00%
	Contar de Geral3	59,38%	1,52%	0,00%	21,95%	11,65%
	Contar de Geral4	7,63%	0,40%	0,00%	3,61%	11,65%
SpaceShuttle	Contar de Geral	10	14	1	17	42
	Contar de Geral2	23,81%	33,33%	2,38%	40,48%	100,00%
	Contar de Geral3	31,25%	21,21%	0,91%	41,46%	16,87%
	Contar de Geral4	4,02%	5,62%	0,40%	6,83%	16,87%
Valentiniana	Contar de Geral	3	18	11	9	41
	Contar de Geral2	7,32%	43,90%	26,83%	21,95%	100,00%
	Contar de Geral3	9,38%	27,27%	10,00%	21,95%	16,47%
	Contar de Geral4	1,20%	7,23%	4,42%	3,61%	16,47%
Total Contar de Geral		32	66	110	41	249
Total Contar de Geral2		12,85%	26,51%	44,18%	16,47%	100,00%
Total Contar de Geral3		100,00%	100,00%	100,00%	100,00%	100,00%
Total Contar de Geral4		12,85%	26,51%	44,18%	16,47%	100,00%

Figura 42 - Tabela de contingências: Modelo por Opinião geral sobre o veículo

Já removemos o campo vazio que existia na variável Modelo (ver Figura 15). Como interpretar a tabela acima? Pelas frequências em cada cruzamento poderemos nos confundir, mas temos três percentuais que podem ser usados. O Contar de Geral2 refere-se ao total das linhas (81, 56, 29, 42 e 41, ou seja, os totais da variável Modelo). O Contar de Geral3 refere-se ao total das colunas (32, 66, 110 e 41, ou seja, os totais da variável Geral). E Contar de Geral4 refere-se ao total da tabela inteira, no caso 249 (que é o número de dados registrados, não vazios).

No presente caso, imagine que quiséssemos avaliar se há diferença entre as opiniões dos clientes da montadora em função do modelo de veículo adquirido. Devemos, então, usar os percentuais por linha, comparando os resultados cada modelo com o Total % por linha. Observe que: do total de clientes, 12,85% tem opinião Bastante satisfeito, 16,47% Satisfeito, 26,51% Insatisfeito, e 44,18% Muito insatisfeito. Se houver diferença de opinião dependendo modelo os

percentuais por linha de cada tipo devem ser substancialmente diferentes (pelo menos 5%) desses valores. Na **Erro! Fonte de referência não encontrada.** é possível constatar:

- os modelos Chiconaultla e DeltaForce3 têm percentuais de “Muito insatisfeito” bem superiores aos dos outros modelos, por exemplo, 85,19% e 51,79% respectivamente, acima do percentual total de muito insatisfeitos (44,18%). Isso significa que, embora mais vendidos, aqueles modelos estão decepcionando seus compradores.

- os veículos LuxuriousCar e SpaceShuttle tem um percentual de Bastante satisfeito acima do geral (65,52% e 23,81%, respectivamente, contra 12,85%).

Conclui-se que há relação entre a opinião geral e o modelo, pois os percentuais por modelo são significativamente diferentes dos percentuais totais de opinião geral.

Se quisermos observar a contribuição individual de cada combinação, devemos usar os percentuais gerais:

- isoladamente, o cruzamento com maior impacto no geral é a combinação Chiconaultla – Muito insatisfeito, que responde por 27,71% das respostas totais (249 entrevistados). O que é bastante preocupante para a futura posição de mercado da montadora.

Da mesma forma que para uma variável qualitativa podemos construir um gráfico de barras, mas aqui múltiplas, para representar a tabela da Figura 42. Usando procedimentos análogos aos da Figura 17 à Figura 21 vamos ter:

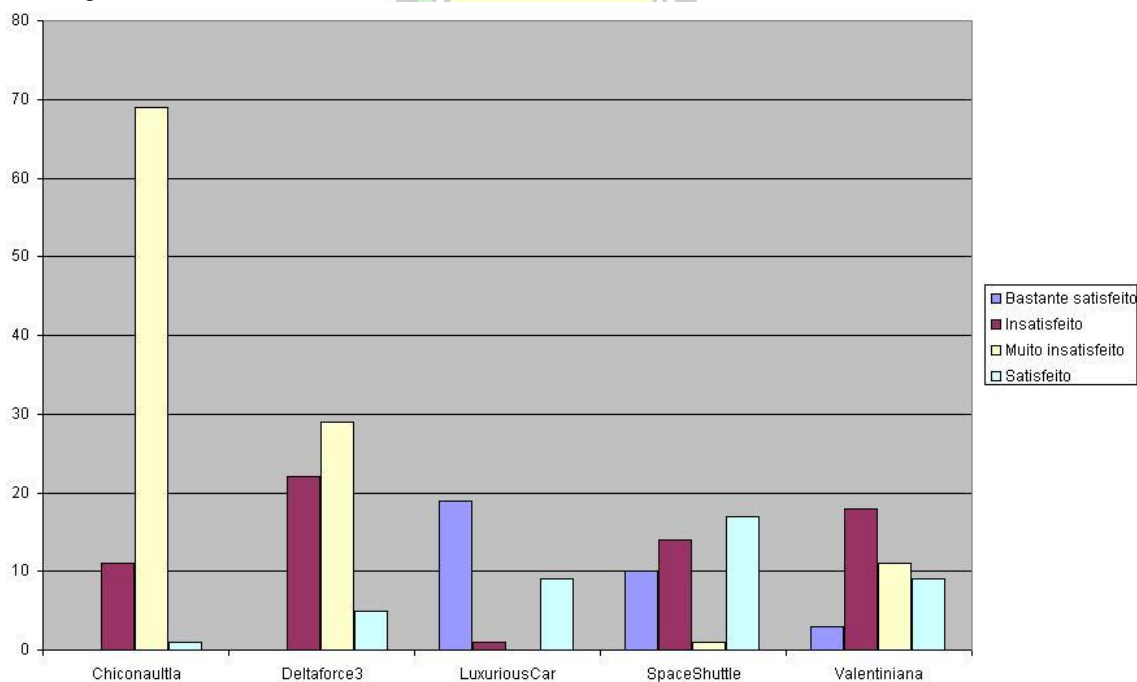


Figura 43 - Gráfico em barras múltiplas da Opinião Geral por Modelo

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).

Podemos transformar o gráfico acima em outro bem interessante, o gráfico em colunas 100% empilhadas, que permitirá visualizar os percentuais por linha graficamente. Basta selecionar o gráfico da Figura 43 e pressionar o botão direito do mouse, o que resulta nas opções mostradas na Figura 44.

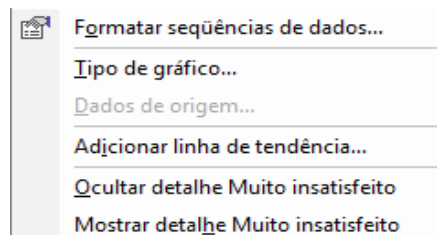


Figura 44 – Opções de gráfico dinâmico

Se pressionarmos “Tipo de gráfico” na Figura 44, chegaremos a uma situação semelhante a da Figura 31. Podemos procurar, entre os tipos de gráficos de colunas, o de colunas 100% empilhadas, como mostrado na Figura 45. E pressionando “OK” na Figura 45 resulta o gráfico mostrado na Figura 46.

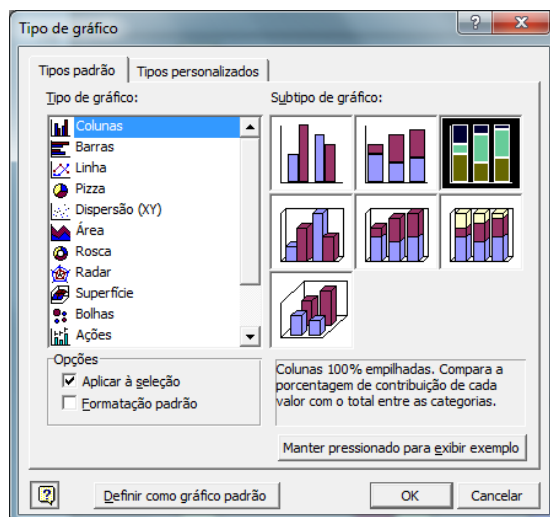


Figura 45 – Opções de gráfico dinâmico – colunas 100% empilhadas

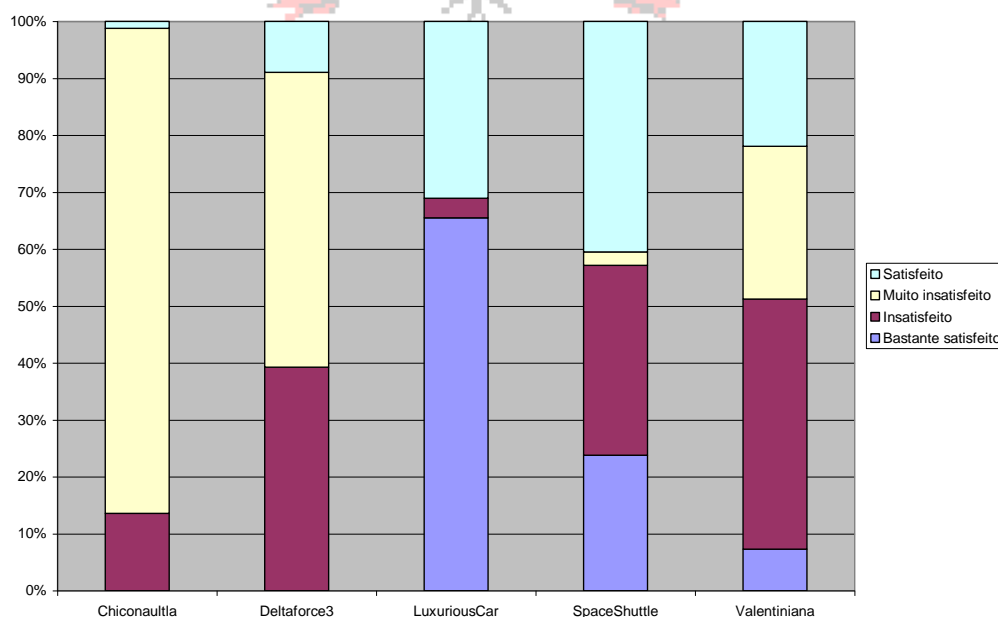


Figura 46 – Gráfico de colunas 100% empilhadas: Opinião geral por Modelo

É possível ver claramente que os percentuais de insatisfeitos e satisfeitos mudam substancialmente de um modelo para outro, exatamente como mostrado na Figura 42 (percentuais por linha).

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.xls”, o procedimento pode ser semelhante ao utilizado para a variável Modelo (seção 1.1): contudo ao construir a tabela dinâmica o Excel poderá selecionar "Soma de Remodelação" como ação (porque os valores da variável são números), e você precisará 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:

Remode	Dados	Total
0	Contar de Remodelação	2
	Contar de Remodelação2	0,80%
1	Contar de Remodelação	57
	Contar de Remodelação2	22,80%
2	Contar de Remodelação	123
	Contar de Remodelação2	49,20%
3	Contar de Remodelação	59
	Contar de Remodelação2	23,60%
4	Contar de Remodelação	9
	Contar de Remodelação2	3,60%
Total Contar de Remodelação		250
Total Contar de Remodelação2		100,00%

Figura 47 - Tabela de frequência: anos de Remodelação

É possível perceber que uma parcela significativa (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 a variável Modelo (seção 1.1) podemos construir um histograma. Os cuidados lá tomados, de remover os percentuais da tabela dinâmica, precisam ser repetidos, mas cabe outra advertência. Imagine que você obtivesse uma tabela como a mostrada na Figura 48.

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 48 - 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 47.

Como a nossa variável anos de Remodelação não tem valores com frequência nula, podemos construir o gráfico diretamente a partir da tabela dinâmica, tal como fizemos para a variável Modelo.

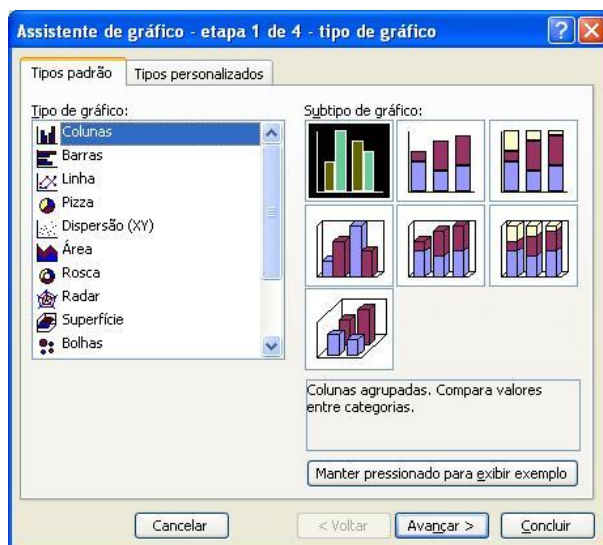


Figura 49 - Assistente gráfico - etapa 1

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 Avançar chegaremos à opção “Dados de origem”. Se, como no nosso caso, a tabela dinâmica foi selecionada diretamente o próprio Excel® preencherá o intervalo de dados. Suponha que a tabela esteja entre as células M2 e O14, obteremos a tela da Figura 50.

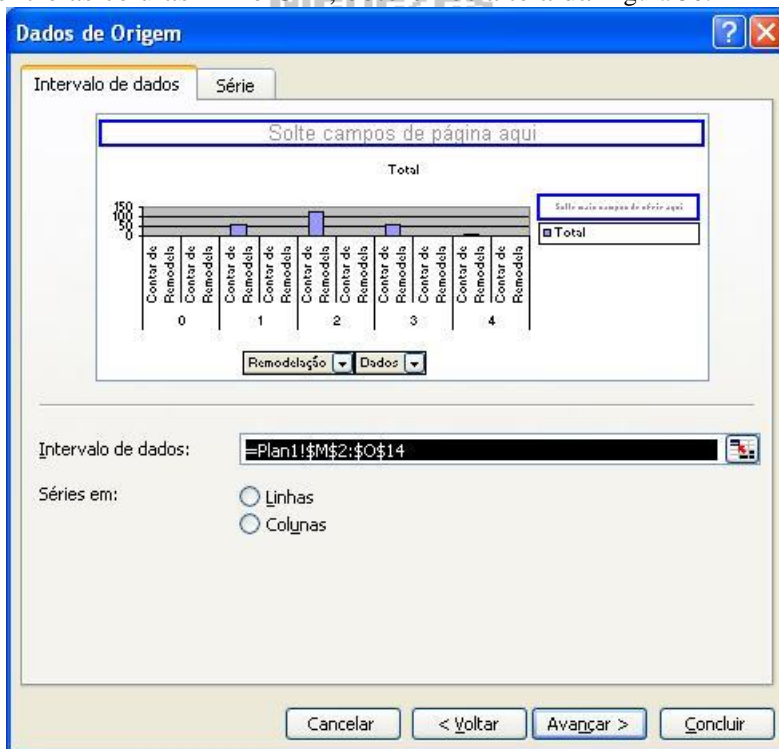


Figura 50 - Assistente gráfico: etapa dos dados de origem

Com os dados selecionados, basta pressionar “Avançar” e chegar à Figura 51.

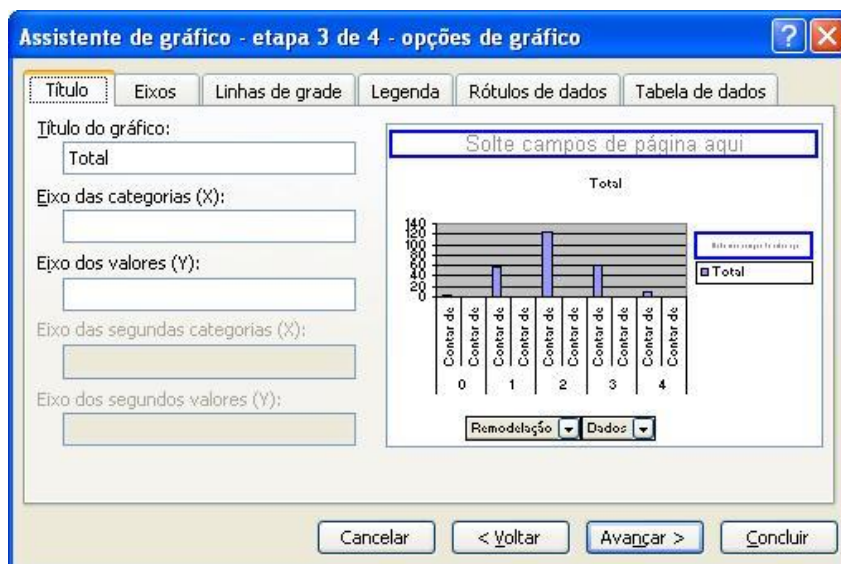


Figura 51 - Assistente gráfico: etapa 3.

Nesta etapa podemos modificar o título dos eixos, o título do próprio gráfico, a eventual colocação de legenda, entre outros aspectos. Posteriormente, basta definir onde o resultado do gráfico será posto, como mostra a Figura 52.

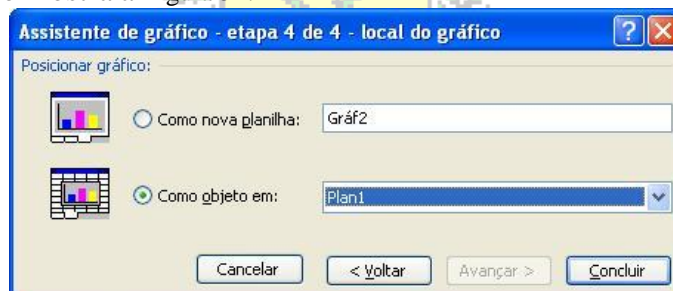


Figura 52 - Assistente gráfico: etapa 4.

Após todas as escolhas chegaremos ao gráfico mostrado na Figura 53.

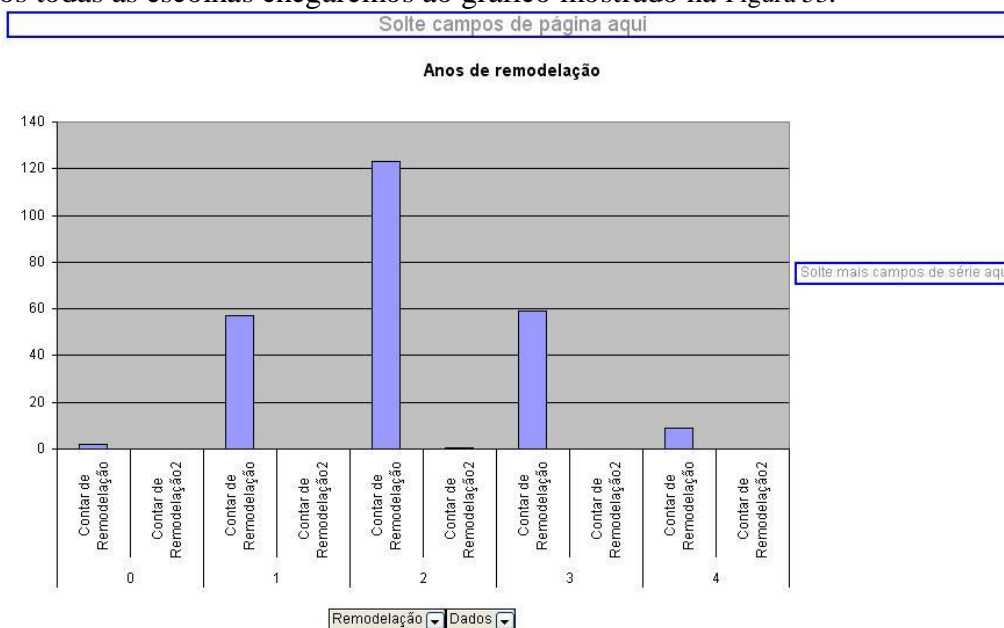


Figura 53 - Gráfico de anos de remodelação - 1a versão

Temos uma situação semelhante a da Figura 18. Podemos aplicar os mesmos procedimentos vistos da Figura 19 à Figura 21, removendo os percentuais e os campos dinâmicos, resultando no gráfico final da Figura 54:

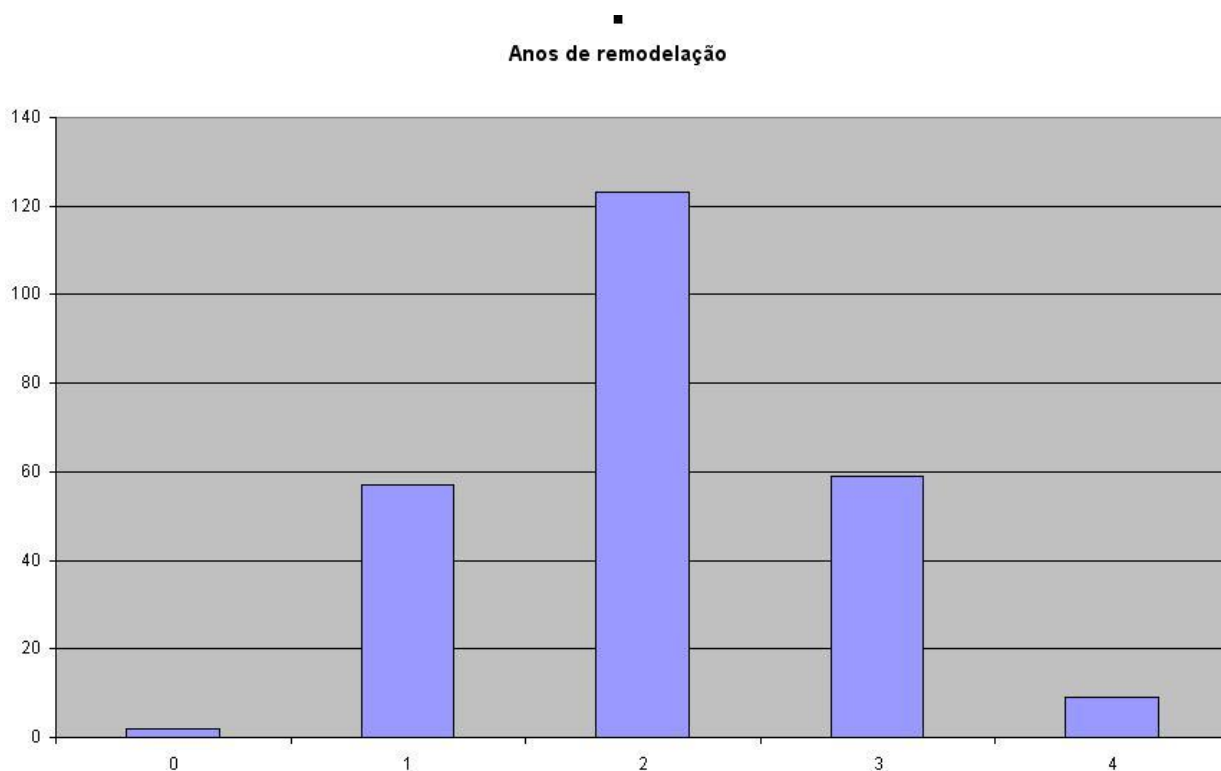


Figura 54 - Gráfico de anos de remodelação - 2a versão

Se você observar o material da Unidade 3 – Análise Exploratória de Dados I – verá que há algo estranho no histograma acima... As barras não são justapostas, pois o Excel ® tem como padrão inserir espaços entre as barras dos gráficos em colunas. Isso precisa ser corrigido. O gráfico da Figura 54 NÃO É um histograma, pois as barras não estão justapostas. Precisamos reduzir a distância entre as barras. Posicione o cursor exatamente sobre uma das barras e pressione o botão esquerdo, resultando na Figura 55.

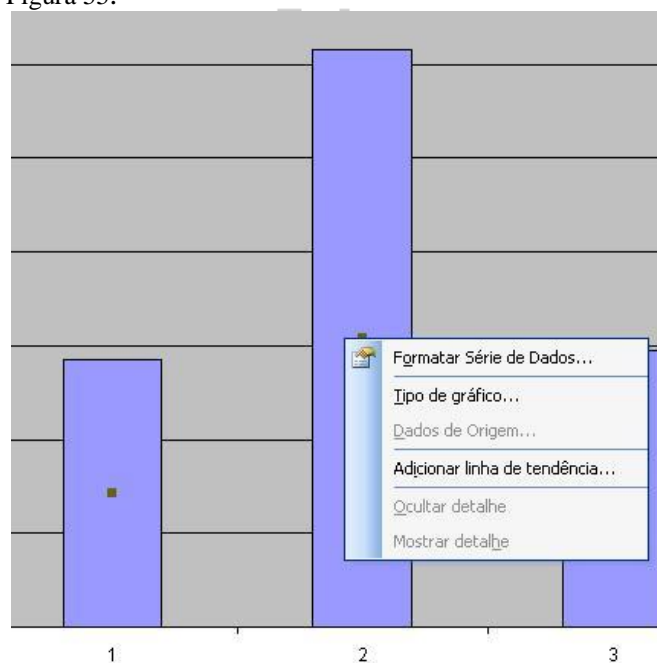


Figura 55 – Opções de gráficos

Ao pressionar “Formatar Série de Dados” chegaremos em várias opções, na Figura 56.



Figura 56 - Opções de formatação da série de dados

Você pode mudar os "Padrões" do gráfico: cores das barras, espessura das linhas, etc. Pode mudar as características de um eixo, dos rótulos de dados, a ordem da sequência, o acréscimo de barras de erros, simplesmente selecionando as opções corretas. Para reduzir as distâncias entre as barras selecionamos "Opções", o que nos levará à tela mostrada na Figura 57.

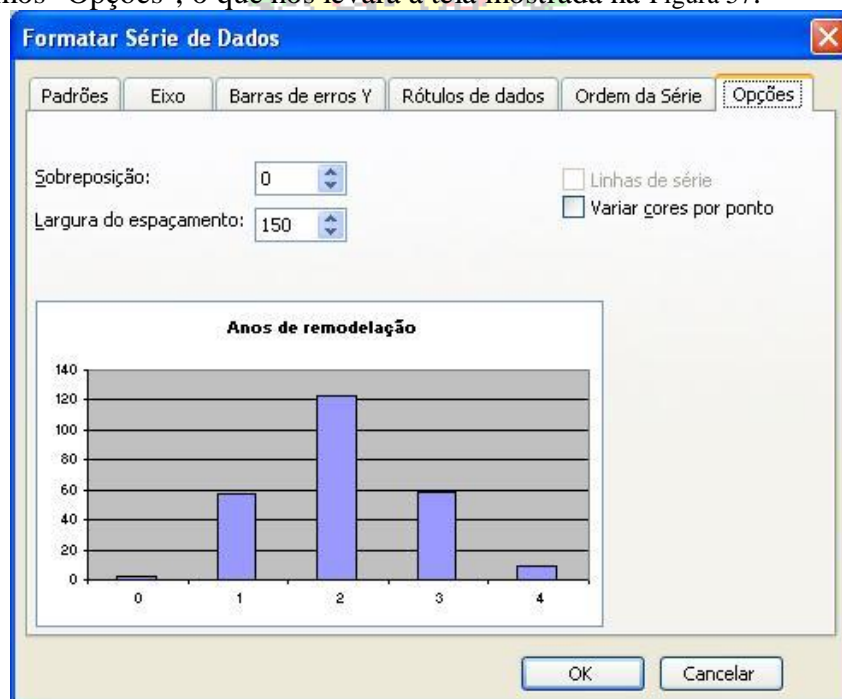


Figura 57 - Opções para as colunas

Na opção "Largura do espaçamento" faça o valor ser zero (o padrão é 150). O resultado final será o histograma mostrado na Figura 58, conceitualmente correto.

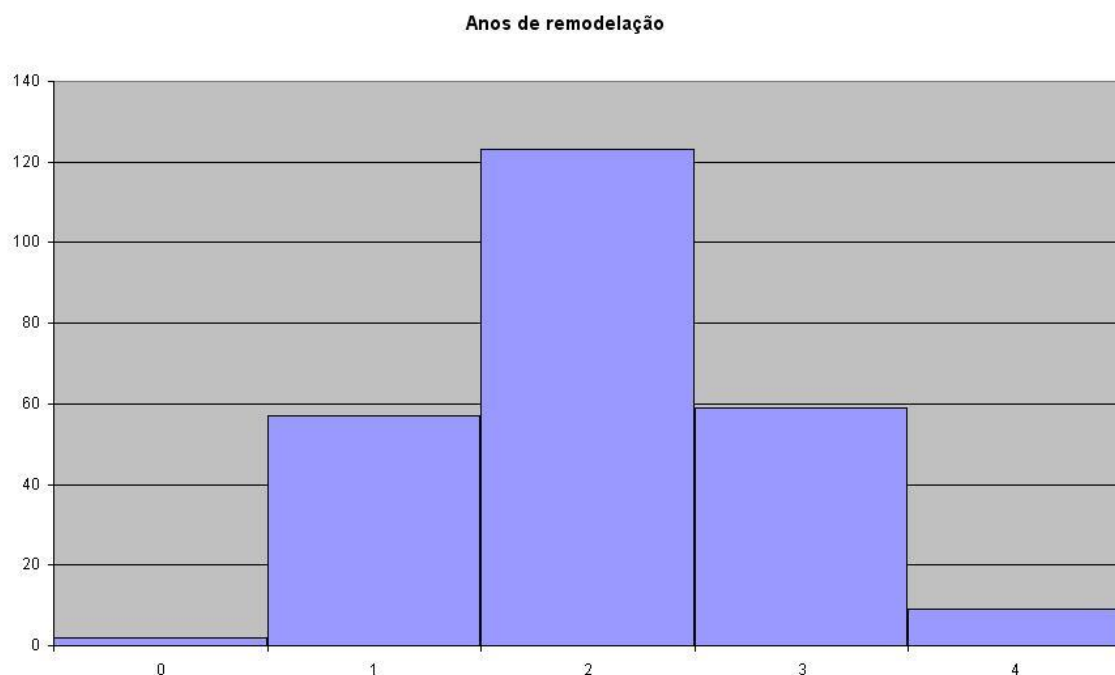


Figura 58 - Histograma de anos de remodelação

Percebe-se claramente que a percepção dos anos de remodelação está concentrada em torno de 2 anos, diminuindo as frequências progressivamente, à medida que se afastam de 2.


2.2 – Procedimentos para variáveis quantitativas contínuas

Se a variável de interesse for quantitativa contínua, como Renda mensal, "Renda" no arquivo "AmostraToyord.xls", 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 Excel: 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 Excel, como "MÁXIMO", "MÍNIMO", "CONT.SE", entre outras.

2.2.1 – Tabela de frequências agrupada em classes

Vamos observar os passos apresentados na Unidade 3:

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. Observe que na barra de ferramentas do Excel há um botão chamado "Colar Função" (), veja a Figura 59:

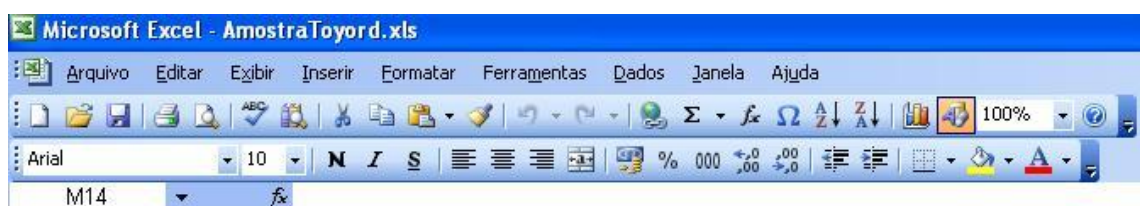


Figura 59 - Barra de ferramentas do Excel: "Colar Função"

Se você pressionar  surgirá a tela vista na Figura 60:

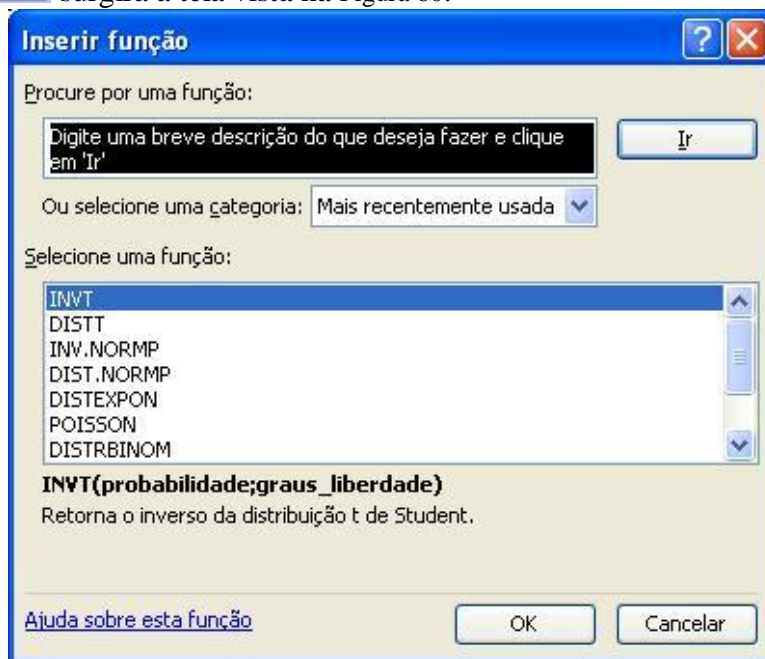


Figura 60 - Funções disponíveis no Excel

Você pode procurar diretamente a função desejada através do primeiro campo no alto da figura. No segundo campo há a descrição das categorias disponíveis de funções: o padrão é “Mais recentemente usada”. Se você clicar na seta à direita do campo verá que há outras opções: matemática, trigonométrica, e Estatística. Selecione esta última e surgirá a Figura 61.

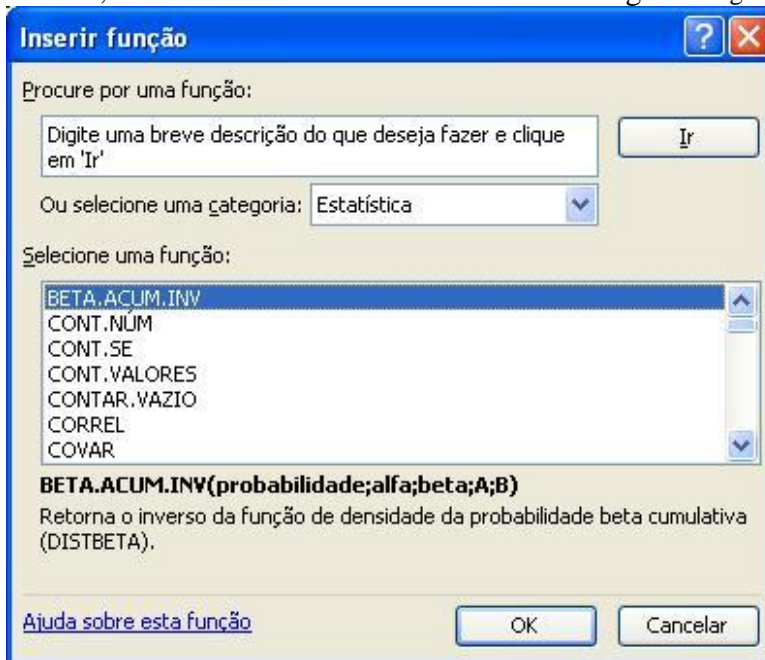


Figura 61 - Funções estatísticas

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 “OK” chegamos à Figura 62:

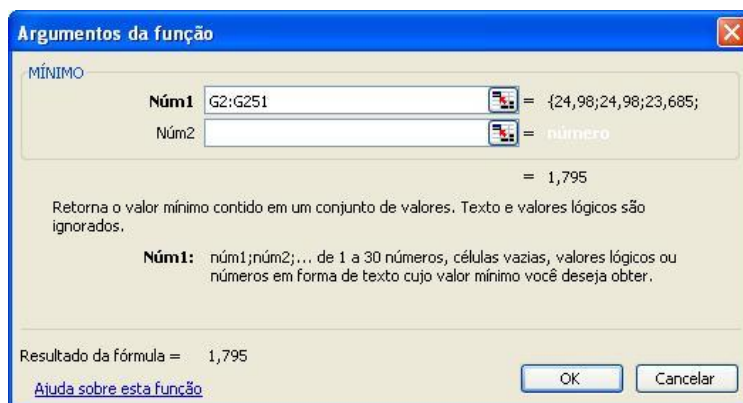


Figura 62 - Função Mínimo

Você precisa escolher o intervalo de valores da variável, para que o Excel 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 sobre esta função”, que pode ser muito útil.

Vamos colocar o resultado na célula O2. O mesmo resultado poderia ser obtido simplesmente digitando a fórmula diretamente na célula: `=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, por exemplo a O3. Para calcular o intervalo podemos colocar uma fórmula em uma célula, por exemplo na O4 fazendo a subtração entre máximo e mínimo. Veja a Figura 63.

	N	O	P		N	O	P		N	O
1				1				1		
2	Mínimo	<code>=MÍNIMO(G2:G251)</code>		2	Mínimo	1,795		2	Mínimo	1,795
3	Máximo	<code>MÍNIMO(...)</code>		3	Máximo	<code>=MÁXIMO(G2:G251)</code>		3	Máximo	86,015
4				4				4	Intervalo	<code>=O3-O2</code>

Figura 63 - Valores mínimo e máximo de Renda

A menor renda mensal dos clientes pesquisados é de 1,795 salários mínimos (cerca de R\$1116,49 em janeiro de 2012), e a maior é de 86,015 salários mínimos (cerca R\$ 53 mil em janeiro de 2012). 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 (veja discussão na Unidade 3). No arquivo “AmostraToyord.xls” há 250 pessoas, extraíndo a raiz quadrada (a função `RAIZ`, na categoria "Matemática e trigonometria" do Excel®), 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. O número de classes e a fórmula para calcular a amplitude das classes poderiam ser colocados nas células O5 e O6, respectivamente, como mostrado na Figura 64. 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.

	N	O
1		
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5	Classes	5
6	Amplitude	<code>=O4/O5</code>

Figura 64 – Amplitude dos valores de Renda

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. Para amplitude de 84,22, e valor inicial igual ao mínimo, com 5 classes, o resultado poderá ser calculado em células apropriadas, como mostrado a partir da Figura 65.

	N	O		N	O
1	Renda		1	Renda	
2	Mínimo	1,795	2	Mínimo	1,795
3	Máximo	86,015	3	Máximo	86,015
4	Intervalo	84,22	4	Intervalo	84,22
5	Classes	5	5	Classes	5
6	Amplitude	16,844	6	Amplitude	16,844
7			7		
8	Classes		8	Classes	
9	Limite inferior	Limite superior	9	Limite inferior	Limite superior
10	=O2		10	1,795	=N10+\$O\$6

Figura 65 – Cálculo dos limites inferior e 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 N10) mais o conteúdo da célula O6 (amplitude das classes). A amplitude das classes é referência absoluta, para possibilitar arrastar a fórmula até o final da tabela. Precisamos também definir os limites da segunda classe da tabela, na Figura 66.

	N	O		N	O
1	Renda		1	Renda	
2	Mínimo	1,795	2	Mínimo	1,795
3	Máximo	86,015	3	Máximo	86,015
4	Intervalo	84,22	4	Intervalo	84,22
5	Classes	5	5	Classes	5
6	Amplitude	16,844	6	Amplitude	16,844
7			7		
8	Classes		8	Classes	
9	Limite inferior	Limite superior	9	Limite inferior	Limite superior
10	1,795	18,639	10	1,795	18,639
11	=O10		11	18,639	=N11+\$O\$6

Figura 66 – 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 N11 e O11 até o fim da tabela e completar as 5 classes. Veja a Figura 67.

	N	O
1	Renda	
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5	Classes	5
6	Amplitude	16,844
7		
8	Classes	
9	Limite inferior	Limite superior
10	1,795	18,639
11	18,639	35,483
12	35,483	52,327
13	52,327	69,171
14	69,171	86,015

Figura 67 – Classes de renda

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 35,483 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.

A implementação é vista a partir da Figura 68.

	N	O	P	Q	R	S	T	U
1	Renda							
2	Mínimo	1,795						
3	Máximo	86,015						
4	Intervalo	84,22						
5	Classes	5						
6	Amplitude	16,844						
7								
8	Classes							
9	Limite inferior	Limite superior	Frequências					
10	1,795	18,639	=CONT.SE(\$G\$2:\$G\$251,"<"&O10)-CONT.SE(\$G\$2:\$G\$251,"<"&N10)					
11	18,639	35,483						
12	35,483	52,327						
13	52,327	69,171						
14	69,171	86,015						

Figura 68 – Obtenção da frequência da primeira classe de renda

Observe que se contam 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 G5001 (valores de Renda) como referência absoluta, para permitir arrastar a fórmula até a célula P14 (ver Figura 69). O símbolo & permite usar como referência o conteúdo de outra célula da planilha. No caso, o limite superior (que está na célula O10) e o inferior (que está na célula N10).

	N	O	P	Q	R	S	T	U
1	Renda							
2	Mínimo	1,795						
3	Máximo	86,015						
4	Intervalo	84,22						
5	Classes	5						
6	Amplitude	16,844						
7								
8	Classes							
9	Limite inferior	Limite superior	Frequências					
10	1,795	18,639	98					
11	18,639	35,483	102					
12	35,483	52,327	38					
13	52,327	69,171	9					
14	69,171	86,015	=CONT.SE(\$G\$2:\$G\$251,"<="&O14)-CONT.SE(\$G\$2:\$G\$251,"<"&N14)					

Figura 69 – Obtenção da frequência da última classe de renda

Observe que na obtenção da frequência da última classe de Renda devemos usar o critério <= 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. Precisamos também definir os pontos médios das classes, que passarão a representá-las, como se as frequências obtidas fossem deles, ver Figura 70.

	N	O	P	Q
1	Renda			
2	Mínimo	1,795		
3	Máximo	86,015		
4	Intervalo	84,22		
5	Classes	5		
6	Amplitude	16,844		
7				
8	Classes			
9	Limite inferior	Limite superior	Frequências	Pontos médios
10	1,795	18,639	98	= (N10+O10)/2
11	18,639	35,483	102	
12	35,483	52,327	38	
13	52,327	69,171	9	
14	69,171	86,015	3	

	N	O	P	Q
1	Renda			
2	Mínimo	1,795		
3	Máximo	86,015		
4	Intervalo	84,22		
5	Classes	5		
6	Amplitude	16,844		
7				
8	Classes			
9	Limite inferior	Limite superior	Frequências	Pontos médios
10	1,795	18,639	98	10,217
11	18,639	35,483	102	27,061
12	35,483	52,327	38	43,905
13	52,327	69,171	9	60,749
14	69,171	86,015	3	77,593

Figura 70 – Obtenção dos pontos médios das classes de renda

Há maior número de clientes com renda mais baixa (até 35,495 salários mínimos, cerca de R\$ 22000 em janeiro de 2012, nem tão baixa assim...). Esta tabela pode ser usada para construir um histograma, de forma semelhante ao que foi visto no item 2.1, sem a necessidade de adicionar rótulos. Selecionando a tabela, escolhendo um gráfico de colunas, e reduzindo o espaçamento entre as barras a zero, entre outros ajustes, vamos obter um histograma tal como o da Figura 71.

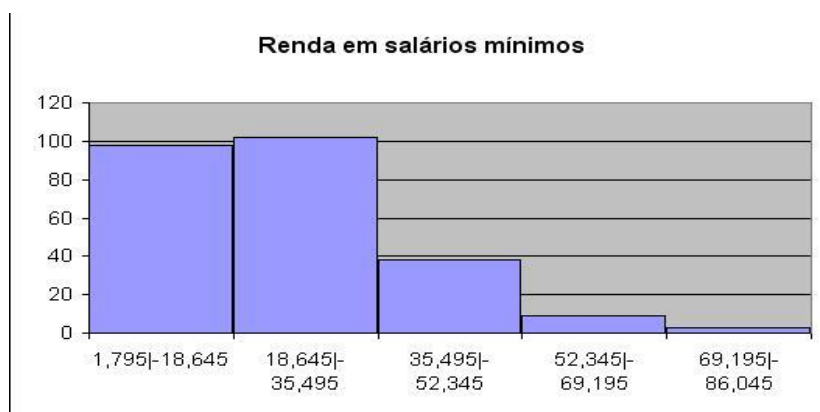


Figura 71 - Histograma das Rendas anuais (agrupados em classes)

2.2.2 - 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.xls”, podemos adotar a seguinte categorização (totalmente arbitrária): os clientes com renda menor do que 2 salários mínimos serão classificados como de renda baixa; aqueles com renda maior ou igual a 2 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 Excel ®. 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.xls”, você poderia pôr o cursor na célula V1. Digite o nome da nova variável, “Renda Categorizada”, e passe o cursor para a célula V2. 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 Excel ®, uma função lógica, a função SE. Podemos inseri-la na célula V2, logo abaixo do nome da nova variável: escolha a categoria “Lógica”, e dentro desta a função SE.

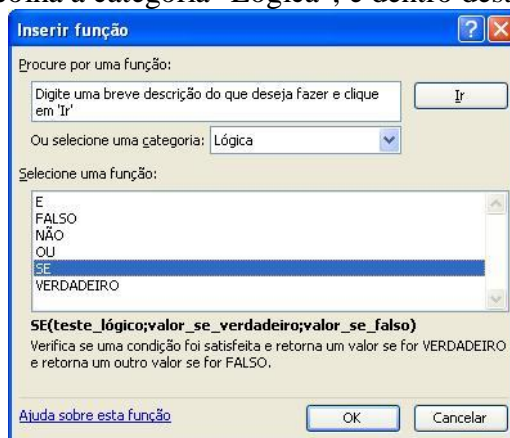


Figura 72 - Função lógica SE

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.

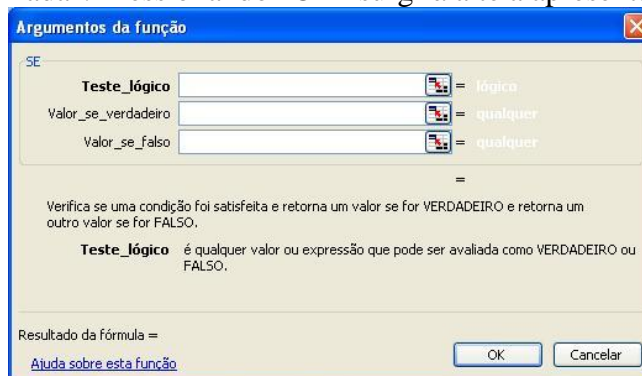


Figura 73 - Argumentos da função SE

A estrutura mostrada na Figura 73 pode ser representada de outra forma, escrevendo diretamente a fórmula na célula. Vamos supor que queremos fazer um teste simples e colocar o resultado na célula V2: se o valor na célula G2 for menor que 2 o valor da célula V2 será "Menor do que 2", caso contrário o valor será "Maior ou igual a 2". A fórmula digitada na célula V2 ficaria (já na sintaxe do Excel ®):

=SE(G2<2;"Menor do que 2";"Maior ou igual a 2")

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.

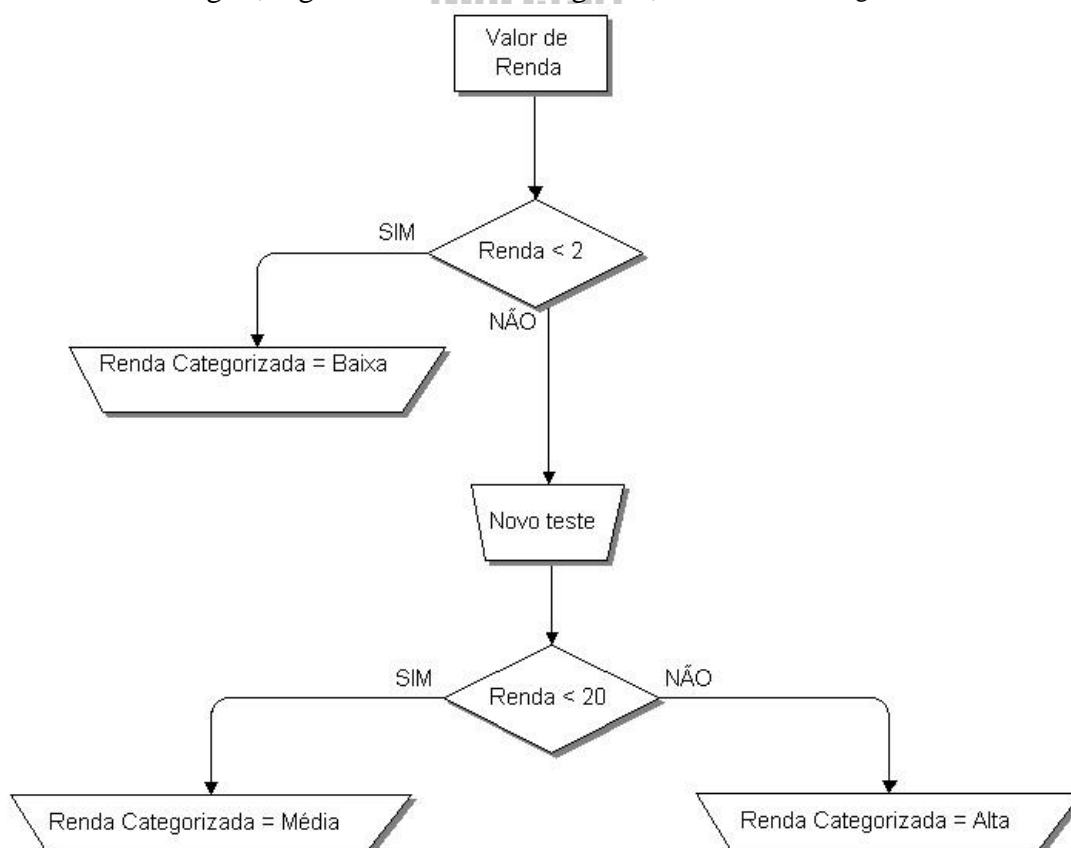


Figura 74 - Procedimento para categorizar a variável Renda

Observe com cuidado a estrutura da função:

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

O procedimento descrito na Figura 74 precisa ser codificado de acordo com a sintaxe do Excel para a função SE, sabendo que o primeiro valor da variável Renda está na célula G2, vamos digitar a fórmula na célula V2, onde deverá estar o primeiro valor da variável Renda Categorizada.

V	W	X	Y
RendaCategorizada			
=SE(G2<2;"Baixa";SE(G2>20;"Alta";"Média"))			

Figura 75 - Categorização da variável Renda no Excel ®

O primeiro teste consiste em verificar se o conteúdo da célula G2 (primeira observação da variável Renda) é menor do que 2. Caso seja, a célula V2 (primeira observação da variável Renda Categorizada) assumirá o valor "Baixa". Caso contrário, se o teste der resultado falso, ao invés de colocarmos um valor que a célula deveria assumir, realizamos mais um teste: inserimos outra função SE para verificar se o valor de G2 é menor do que 20. Caso seja, a célula V2 assumirá o valor "Média". Se essa condição for satisfeita a célula V2 assumirá o valor "Alta", pois se todos os testes anteriores tiveram resultados "FALSO", este é o único valor remanescente. Observe que para cada parênteses aberto precisa haver um fechado, caso contrário o Excel ® acusará erro na fórmula.

Uma vez completa a fórmula na célula V2 podemos "arrastá-la" para baixo, chegando ao final do arquivo, na célula V251. O resultado será uma nova variável, relacionada à Renda, mas categorizada (qualitativa), veja a Figura 76.

V
RendaCategorizada
Alta
Alta
Alta
Média
Média
Alta
Alta
Alta
Alta

Figura 76 - Variável Renda Categorizada

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

X	Y
Contar de RendaCategorizada	
RendaCategorizada	Total
Alta	144
Baixa	1
Média	105
Total geral	250

Figura 77 - Tabela dinâmica da variável Renda Categorizada

Talvez a categorização escolhida não tenha sido a melhor... Observe que na Renda baixa há apenas uma observação.

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 Excel ® 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 Excel.

Quadro 1 - Medidas de síntese e funções do Excel

MEDIDAS DE SÍNTESE	FUNÇÕES DO EXCEL
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 Excel realizará o cálculo. Para os dados presentes no arquivo "AmostraToyord.xls", 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 78 - Medidas de síntese da variável Renda

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 calcular medidas como média, mediana, desvio padrão e quartis, é possível proceder como na seção 2.2.1, a partir da Figura 60. Por exemplo, para o caso da média, pressionando o botão “inserir função”, e escolhendo funções estatísticas, podemos escolher média, como na Figura 79:

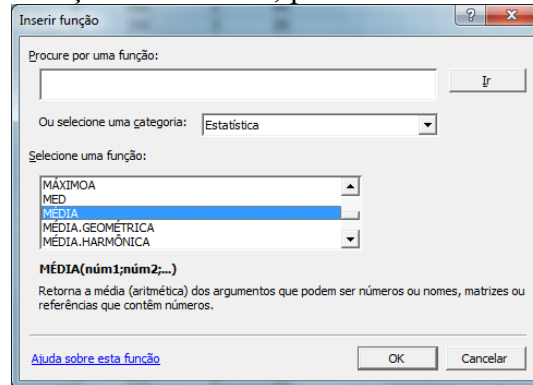


Figura 79 – Inserção de função: Média

Ao pressionar OK na Figura 79 é possível informar o intervalo de células e obter o valor da média. As outras medidas podem ser obtidas de forma análoga.

2.3.2 - Cálculo de Medidas de Síntese utilizando tabelas dinâmicas

No item 1.1, Figura 8, 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.

Vamos imaginar que há interesse em avaliar como está a distribuição da variável "Renda" em função da variável "Modelo", no arquivo “AmostraToyord.xls”. É extremamente interessante conhecer o perfil econômico dos clientes por modelo. O procedimento para construção da tabela dinâmica é semelhante ao mostrado no item 1.2, o layout da tabela será como na Figura 80.

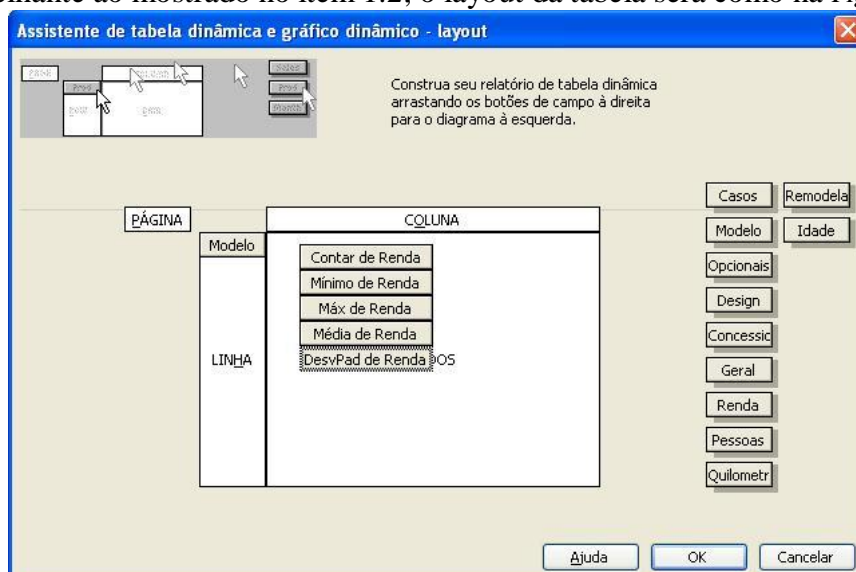


Figura 80 - Tabela dinâmica: layout de medidas de síntese de Renda por Modelo

Observe que no campo "LINHA" colocamos a variável independente "Modelo". NÃO COLOCAMOS nenhuma variável no campo "COLUNA", pois o Excel ® tentará fazer o cruzamento entre linhas em colunas, resultando em uma tabela com uma grande quantidade de células, porque a variável "Renda" é quantitativa contínua. No campo "DADOS" colocamos a variável "Renda", "arrastando-a" da lista de variáveis à direita: a ação padrão do Excel ® para variável quantitativa é "Soma", o que não nos interessa. Se pressionarmos duas vezes "Soma de Renda", chegará às várias opções de medida, mostradas na Figura 8 (no item 1.2): selecione "Contagem de valores" (isso permitirá saber quantas pessoas estão em cada categoria). "Arraste" "Renda" novamente para o campo DADOS, pressione duas vezes "Soma de Renda" e selecione a opção "Média". Repita o procedimento para selecionar "DesvPad" (desvio padrão amostral), "Mínimo" e "Máximo". Observe que não há opção de medidas como mediana, moda e quartis.

Uma vez tendo terminado o layout da tabela, simplesmente concluímos o procedimento (NÃO mudamos as opções da tabela), resultando na Figura 81.

Modelo	Dados	Total
Chiconaultla	Contar de Renda	81
	Mínimo de Renda	1,795
	Máx de Renda	40,16
	Média de Renda	12,7045679
	DesvPad de Renda	6,037939953
Deltaforce3	Contar de Renda	56
	Mínimo de Renda	10,82
	Máx de Renda	48,22
	Média de Renda	22,06357143
	DesvPad de Renda	6,956100182
LuxuriousCar	Contar de Renda	29
	Mínimo de Renda	29,8
	Máx de Renda	86,015
	Média de Renda	50,93189655
	DesvPad de Renda	14,92196388
SpaceShuttle	Contar de Renda	42
	Mínimo de Renda	18,865
	Máx de Renda	47,3
	Média de Renda	33,05035714
	DesvPad de Renda	7,61951005
Valentiniana	Contar de Renda	41
	Mínimo de Renda	13,055
	Máx de Renda	65,39
	Média de Renda	27,35317073
	DesvPad de Renda	8,382676538
Total Contar de Renda		249
Total Mínimo de Renda		1,795
Total Máx de Renda		86,015
Total Média de Renda		25,10542169
Total DesvPad de Renda		14,50520893

Figura 81 - Renda em função do Modelo

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) e a maior renda de todo o conjunto (86,015 salários mínimos). Por outro lado, os compradores de Chiconaultla e DeltaForce3 têm as menores médias e menores mínimos de renda.

Observe nas cinco últimas linhas da tabela as medidas para o conjunto como um todo, exatamente iguais às mostradas na Figura 78, mas sem a mediana e os quartis, o que pode prejudicar um pouco a análise, especialmente se valores discrepantes estiverem presentes nos dados. Não

obstante, trata-se de uma maneira simples de obter medidas em função dos valores de outra variável, e com média, desvio padrão, mínimo e máximo é plenamente possível realizar uma avaliação dos dados.

Poderíamos incluir mais uma variável independente, por exemplo, "Geral", e observar a distribuição da Renda. Apenas tome cuidado: quanto mais variáveis, mais cuidadosa precisa ser a análise dos resultados.

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.xls”, 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 Excel, resultando na figura 2. Selecionando o gráfico Dispersão (XY), obtemos a Figura 82.

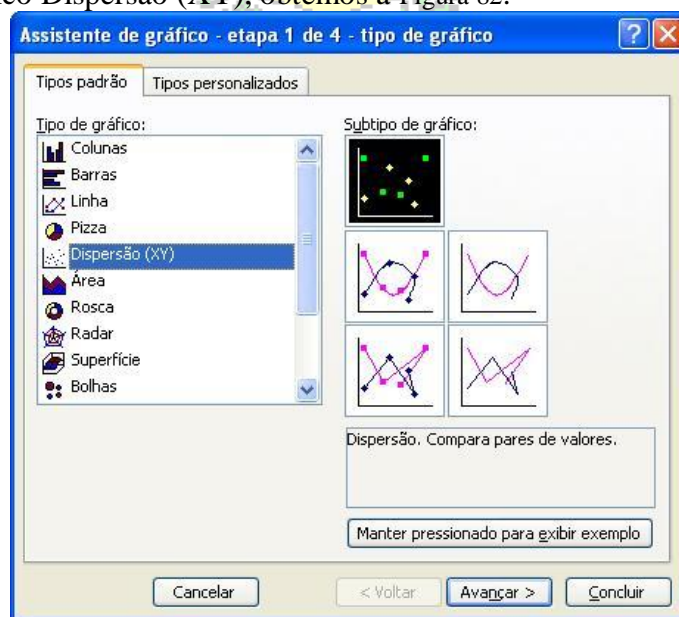


Figura 82 - Assistente gráfico - 1a etapa

Para os nossos interesses o subtipo mais interessante é o padrão, marcado em preto na Figura 82. Pressionando "Avançar" chegaremos a uma tela semelhante à Figura 83.

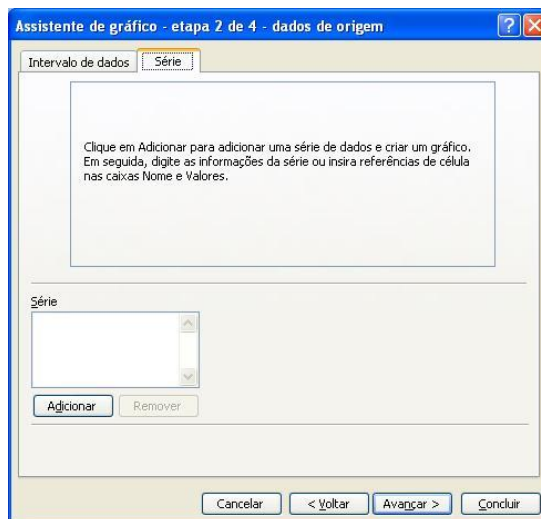


Figura 83- Assistente gráfico - 2a etapa

Em alguns casos o Excel automaticamente adiciona as sequências de dados necessárias para criar o gráfico. Muitas vezes estas sequências incluem dados que não nos interessam. Se isso ocorrer, pressione "Remover" até que todas as sequências sejam retiradas, resultando na tela mostrada na Figura 84. Agora podemos adicionar as sequências de dados de interesse, pressionando "Adicionar", o que resultará na Figura 85.

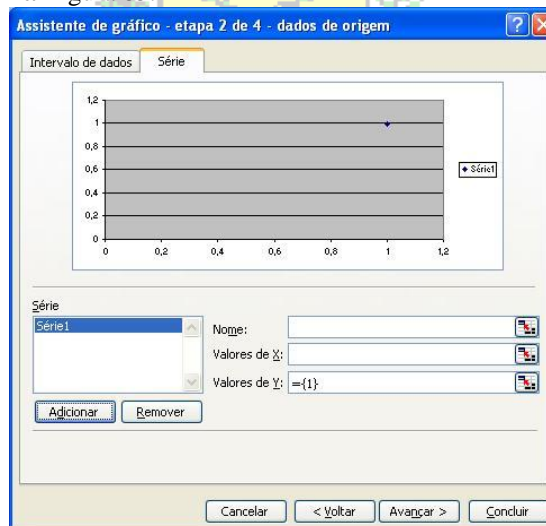


Figura 84 - Assistente gráfico: adição de sequências - início

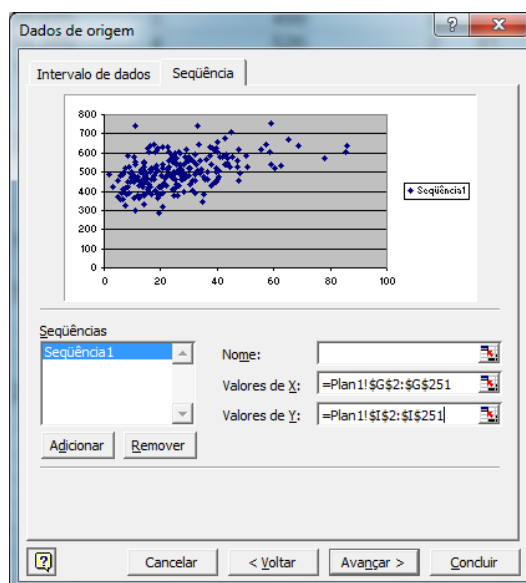


Figura 85 - Assistente gráfico: adição de sequências - completo

Precisamos adicionar os valores de X e de Y (não há necessidade de adicionar valores em "Nome"). Podemos fazer isso de duas formas: ou digitando as referências das células (em "Valores de X" teríamos G2:G251; em "Valores de Y" teríamos I2:I251), ou marcando as células na planilha (pressionando a seta vermelha na extrema direita de cada janela, e marcando as células de interesse na planilha). Após a adição dos dados, o resultado será uma tela semelhante à da Figura 86.

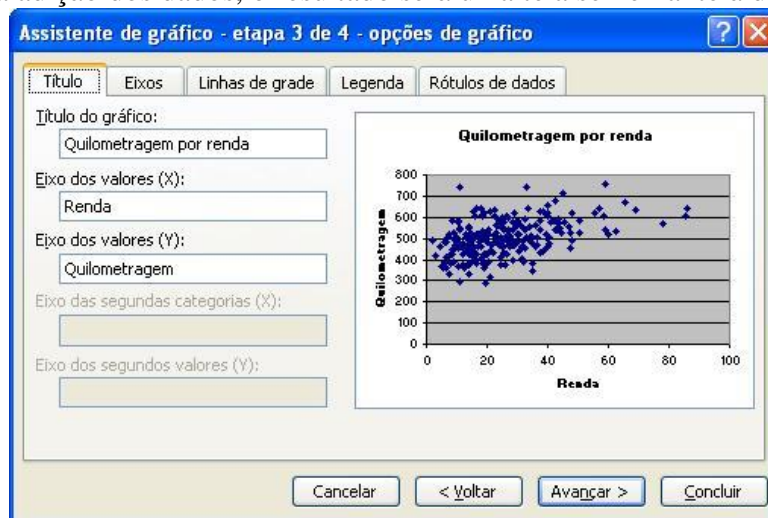


Figura 86 - Assistente gráfico - 3a etapa

Observe que já é possível ter uma ideia do diagrama de dispersão: os dados parecem distribuir-se de forma crescente, com os valores de X começando acima de 0, e os valores de Y variando de 300 até quase 800. Possivelmente teremos que modificar a escala do eixo Y, para que a visualização do gráfico seja mais apropriada: da forma como está o gráfico, os dados estão muito agrupados, o que pode dificultar a análise do diagrama de dispersão.

É necessário pôr um título no gráfico, e identificar as variáveis em cada eixo, incluindo suas unidades. Título: Quilometragem por Renda. Eixo X: Renda. Eixo Y: Quilometragem. Retiramos a legenda, pois não há necessidade neste gráfico. Ao pressionar "Avançar" chegamos na 4ª etapa do assistente, onde escolhemos onde posicionar o gráfico. Independente da escolha o diagrama resultante será:

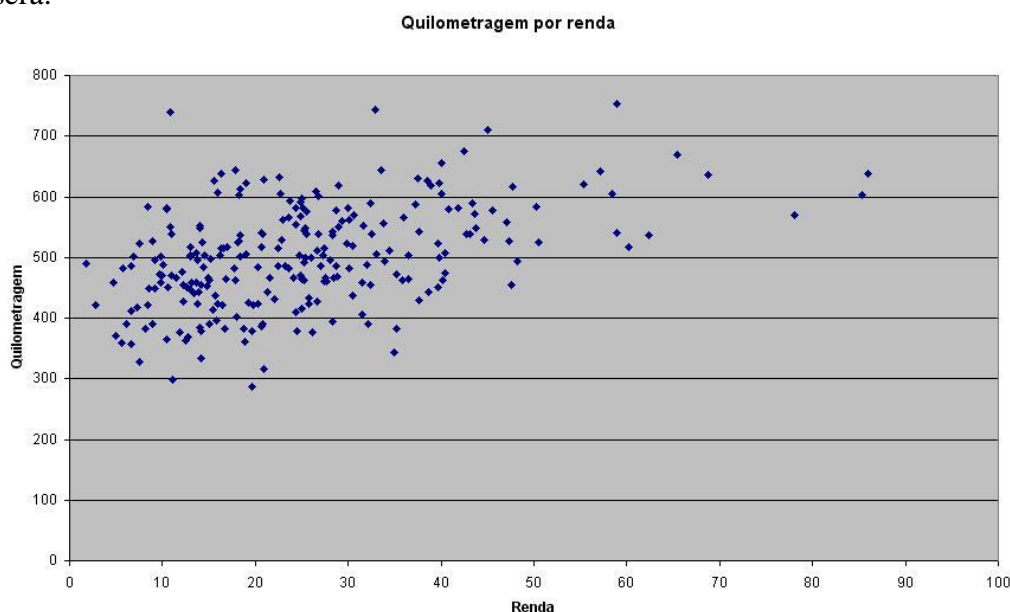


Figura 87 - Diagrama de dispersão de quilometragem por renda

Temos que modificar a escala do eixo Y. Para mudar a escala do eixo Y é preciso colocar o cursor exatamente sobre o eixo Y, e pressionando o botão esquerdo do mouse teremos uma situação como a exposta na Figura 88.

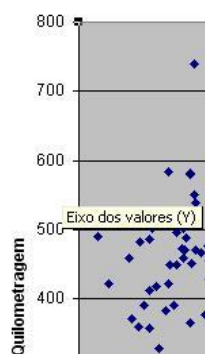


Figura 88 - Seleção do eixo Y

Posteriormente, mantendo o cursor sobre o eixo e pressionando o botão direito do mouse vamos ter acesso às opções relativas ao eixo X, como mostrado na Figura 89.

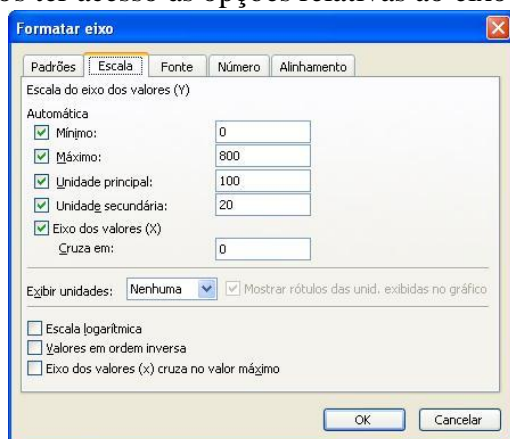


Figura 89 - Opções de formatação de eixo: escala

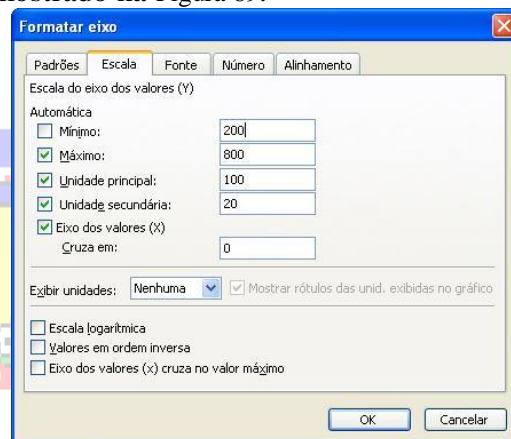


Figura 90 - Formatação de eixo: escala modificada

O comportamento padrão do Excel 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 90. Pressionando "OK" vamos chegar ao gráfico mostrado na Figura 91.

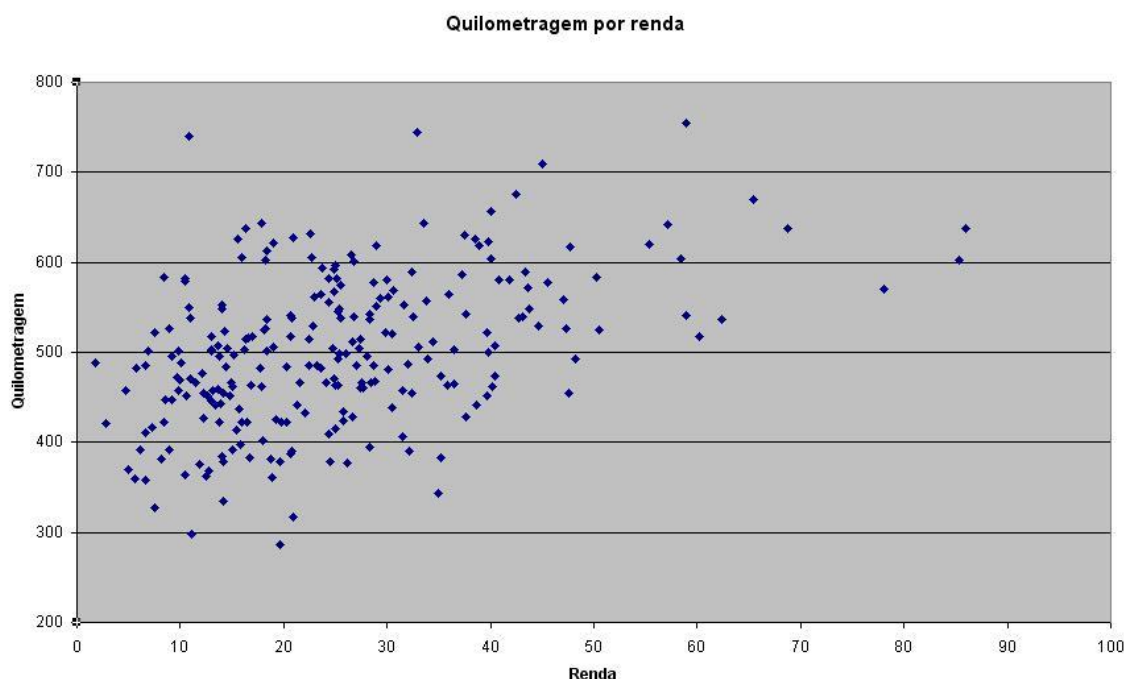


Figura 91 - Diagrama de dispersão de Quilometragem por Renda

Agora podemos fazer uma análise do diagrama de dispersão:

- a força da correlação não é muito grande, pois os pontos estão relativamente distantes entre si.
 - a correlação parece ser positiva, pois se observa que a nuvem de pontos tem um comportamento crescente, ou seja, maiores rendas, maiores quilometragens (e é razoável imaginar que realmente um aumento na renda permita gastar mais com combustível e, portanto, rodar uma maior quilometragem).
 - 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.

