

Procedimentos de Análise Exploratória de Dados utilizando o Excel 2007

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 2007. 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.

Índice Analítico

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)	16
2. Procedimentos para variáveis quantitativas	23
2.1 - Procedimentos para variáveis quantitativas discretas	23
2.2 - Procedimentos para variáveis quantitativas contínuas	26
2.2.1 – Tabela de frequências agrupada em classes	26
2.2.2 – Categorização de uma variável quantitativa contínua	29
2.3 – Cálculo de medidas de síntese	31
2.3.1 – Medidas de síntese para uma variável quantitativa	31
2.3.2 – Medidas de síntese para uma variável quantitativa em função de uma qualitativa – por tabela dinâmica	36
2.3 – Procedimentos para duas variáveis quantitativas (diagrama de dispersão)	39

Índice de Figuras

Figura 1 – Menu de inserção	Figura 2 - Seleção de dados de tabela dinâmica	4
Figura 3 - Tela de seleção de intervalo de dados - minimizada		4
Figura 4 – Layout da Tabela dinâmica		5
Figura 5 – Tabela dinâmica por modelo – apenas frequência		5
Figura 6 – Tabela dinâmica de modelo		6
Figura 7 – Opções do campo Valor		6
Figura 8 – Configuração do Campo Valor: Resumir por		6
Figura 9 – Configuração do Campo Valor: Mostrar valores como		7
Figura 10 – Configuração do Campo Valor: Mostrar valores como % do total		7
Figura 11 – Tabela dinâmica de Modelo: frequências absolutas e relativas – 1ª versão		7
Figura 12 – Tabela dinâmica de Modelo: frequências absolutas e relativas – 2ª versão		8
Figura 13 – Opções de exibição da tabela de Modelo: com remoção das células vazias		8
Figura 14 – Opções de gráficos		8
Figura 15 - Subtipos de gráficos de colunas		9
Figura 16 - Gráfico de colunas de Modelo - 1a versão		9
Figura 17 - Campos do gráfico dinâmico	Figura 18 - Opções do campo %	10
Figura 19 - Gráfico de colunas de Tipo - 2a versão		10
Figura 20 - Opções de design de gráfico dinâmico de colunas		10
Figura 21 - Opções de layout de gráfico dinâmico de colunas		10
Figura 22 - Opções de legenda do gráfico dinâmico		11
Figura 23 - Opções de Título do Eixo Vertical Principal em Gráfico Dinâmico		11
Figura 24 - Gráfico de colunas de Modelo de veículo - Final		12
Figura 25 - Opções de formatação de gráfico e eixo		12
Figura 26 - Opções de Eixo	Figura 27 - Mudança de escala no eixo vertical	13
Figura 28 - Gráfico de colunas de Tipo de serviço – escala vertical modificada		13
Figura 29 - Subtipos de gráficos em Pizza		14
Figura 30 - Gráfico em pizza de Modelo de veículo - 1a versão		14
Figura 31 - Opções de design de gráfico dinâmico em Pizza		14
Figura 32 - Opções de layout de gráfico dinâmico em Pizza		14
Figura 33 - Opções de Rótulo de Dados para Gráfico em Pizza		15
Figura 34 - Mais opções de Rótulo de Dados		15
Figura 35 - Gráfico em pizza de Modelo de Veículo - Final		15
Figura 36 – Célula de destino de uma tabela dinâmica		16
Figura 37 - Tabela dinâmica de Modelo e Geral – Layout inicial		16
Figura 38 - Tabela dinâmica de Modelo e Geral – Layout com frequências absolutas e todos os %		17
Figura 39 - Tabela dinâmica de Tipo x Opinião – Layout com frequências absolutas e todos os %, modificado		17
Figura 40 - % por linha, coluna e total		18
Figura 41 - Tabela dinâmica de Modelo x Geral (opinião geral) (sem células vazias)		18
Figura 42 - Tabela dinâmica Modelo x Geral - opções de formatação		19
Figura 43 - Tabela dinâmica Modelo x Geral - mudança de ordem de valores		19
Figura 44 - Tabela dinâmica de Modelo x Geral		20
Figura 45 - Gráfico em barras múltiplas da Opinião Geral por Modelo		21
Figura 46 – Opções de gráfico dinâmico		22

Figura 47 – Opções de gráfico dinâmico – colunas 100% empilhadas.....	22
Figura 48 – Gráfico de colunas 100% empilhadas: Opinião geral por Modelo.....	22
Figura 49 - Tabela de frequência: anos de Remodelação.....	23
Figura 50 - Tabela de frequência: Anos de educação.....	23
Figura 51 – Opções de gráficos.....	24
Figura 52 - Histograma de Anos de remodelação - 1a versão.....	24
Figura 53 - Gráfico de anos de remodelação - 2a versão.....	24
Figura 54 - Opções de formatação de séries de dados - histograma anos de remodelação	25
Figura 55 - Opções de formatar séries de dados.....	25
Figura 56 - Histograma de anos de remodelação.	25
Figura 57 - Mínimo e máximo para a variável Renda	26
Figura 58 - Intervalo para Renda.....	26
Figura 59 - Amplitude de classes de Renda.....	26
Figura 60 – Cálculo dos limites inferior e do limite superior da primeira classe de Renda	27
Figura 61 – Cálculo dos limites inferior e superior da segunda classe de Renda	27
Figura 62 - Classes de Renda.....	27
Figura 63 – Obtenção da frequência da primeira classe de renda.....	28
Figura 64 - Obtenção da frequência da última classe de Renda	28
Figura 65 – Obtenção dos pontos médios das classes de renda.....	28
Figura 66 - Histograma das Rendas anuais (agrupados em classes)	29
Figura 67 - Procedimento para categorizar a variável Renda.....	30
Figura 75 - Categorização da variável Renda no Excel ®.....	30
Figura 69 - Variável Renda Categorizada.....	31
Figura 70 - Tabela dinâmica da variável Renda Categorizada	31
Figura 71 - Fórmulas no Excel 2007	31
Figura 72 - Fórmulas no Excel 2007 – Mais funções	32
Figura 73 - Medidas de síntese de Renda - 1a parte	32
Figura 74 - Funções disponíveis no Excel	32
Figura 75 - Funções estatísticas no Excel - 1a abordagem.....	33
Figura 76 - Argumentos da função Média (para Renda) - 1a versão.....	33
Figura 77 - Argumentos das funções Mediana e Moda (Renda) - 1a versão.....	33
Figura 78 - Argumentos da função desvio padrão (Renda) - 1a versão.....	34
Figura 79 - Argumentos das funções Quartil Inferior e Superior (Renda) - 1a versão	34
Figura 80 - Argumentos das funções Mínimo e Máximo (Renda) - 1a versão.....	34
Figura 81 - Medidas de síntese de Renda	34
Figura 82 - Cálculo do CV%	35
Figura 83 - Cálculo das diferenças entre os quartis e a mediana	35
Figura 84 - Cálculo dos limites para valores discrepantes	35
Figura 85 - Medidas de síntese para Renda (completas).....	35
Figura 86 - Layout da tabela dinâmica Renda x Modelo 1a parte	37
Figura 87 - Layout da tabela dinâmica Renda x Modelo 2a parte	37
Figura 88 – Formas de resumir os dados de Renda	38
Figura 89 - Layout da tabela dinâmica Renda x Tipo 3a parte.....	38
Figura 90 - Tabela dinâmica de Renda por Tipo.....	39
Figura 91 – Inserção de gráfico de dispersão - 1a etapa	40
Figura 92- Inserção de gráfico de dispersão - 2a etapa.....	40
Figura 93 - Adição de sequências - início.....	40
Figura 94 - Adição de sequências – X e Y.....	41
Figura 95 – Adição de sequências - dados inseridos	41
Figura 96 - Diagrama de dispersão de quilometragem por renda – 1 versão	41
Figura 97 - Diagrama de dispersão de Quilometragem por Renda.....	42

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)

Independente se as variáveis de interesse são qualitativas ou quantitativas o procedimento para construção da tabela dinâmica é o mesmo. Posicione o cursor em algum ponto do conjunto de dados, e procure no menu Inserir por Tabela Dinâmica, como na Figura 1.

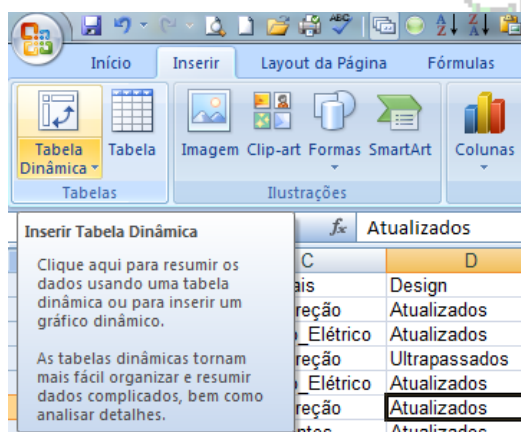


Figura 1 – Menu de inserção

Pressionando Tabela Dinâmica o Excel 2007 apresenta o resultado da Figura 1.

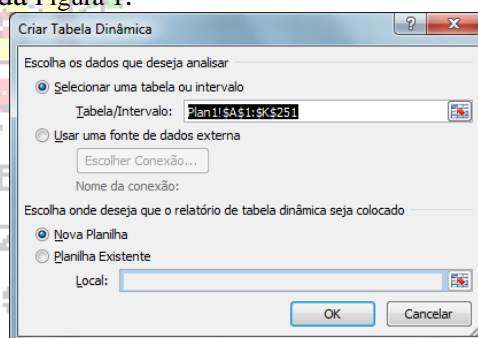


Figura 2 - Seleção de dados de tabela dinâmica

Ao pôr o cursor em alguma parte do banco de dados o Excel 2007 seleciona automaticamente todas as variáveis adjacentes: no presente caso todas as células compreendidas entre A1 e K251: da variável Caso até a Idade. Pode-se escolher, eventualmente, uma fonte externa ou dados de outra planilha, ou outro intervalo de dados na mesma planilha. Para fazer isso basta pressionar o botão com seta vermelha na linha “Tabela/Intervalo” da Figura 1, resultando na Figura 3:

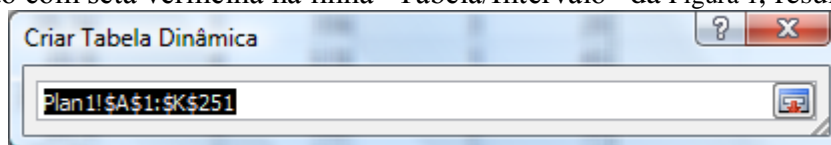


Figura 3 - Tela de seleção de intervalo de dados - minimizada

Com a seleção neste formato você pode procurar o intervalo desejado e selecioná-lo utilizando o mouse. Para retornar à situação da Figura 1 basta pressionar a pequena seta vermelha novamente. Por conveniência, vamos pôr a tabela dinâmica em uma nova planilha. Então, pressionando OK, chega-se à **Erro! Fonte de referência não encontrada..**

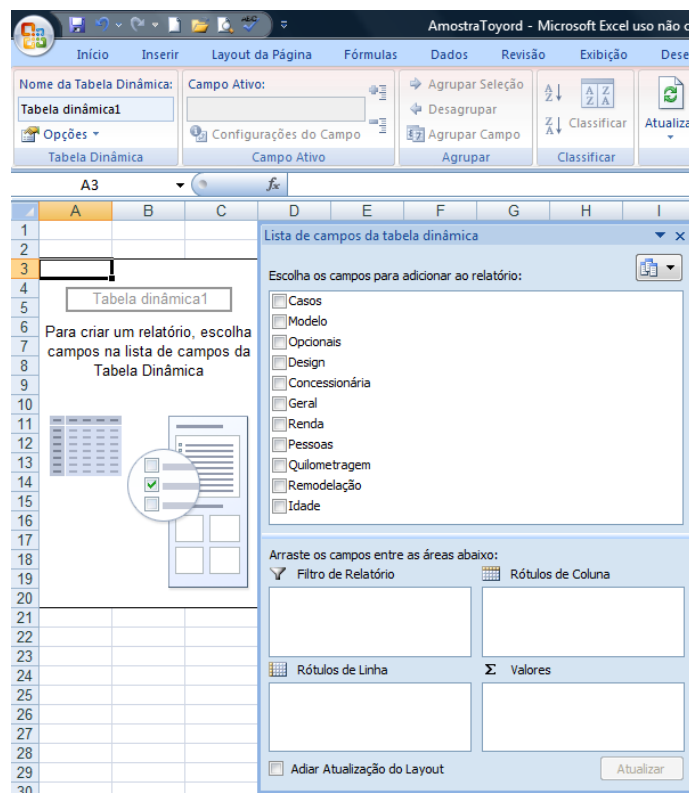


Figura 4 – Layout da Tabela dinâmica

A versão 2007 do Excel automaticamente cria a tabela, vazia, com os campos disponíveis para adição à direita. Observe à direita os nomes das variáveis existentes no arquivo de dados, e que foram selecionadas na etapa anterior. Vamos imaginar que há interesse apenas na variável Modelo, devemos selecioná-la e arrastá-la até o campo "Rótulos de Linha", ou "Rótulos de Coluna". As outras variáveis não farão parte da tabela. Devemos também arrastar a variável Modelo para o campo "Valores". O resultado pode ser visto na Figura 5.

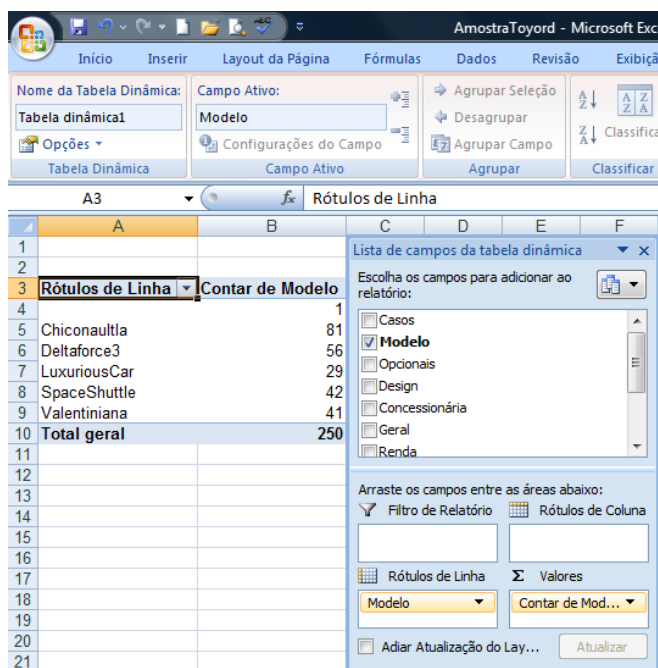


Figura 5 – Tabela dinâmica por modelo – apenas frequência

Observe que a tabela contém apenas as frequências absolutas de cada valor, além das células vazias. Mas, usualmente é imprescindível apresentar as frequências relativas também, para que seja possível fazer comparações com outros conjuntos de dados. Para tanto devemos arrastar novamente Modelo para o campo Valores, o que nos levará à Figura 6: as frequências absolutas dos valores serão apresentadas mais uma vez, e precisamos modificá-las para que a tabela contenha as frequências absolutas E as relativas. Para fazer isso, devemos pressionar a seta▼ do segundo botão do campo Valores, o que leva às opções da Figura 7.

	A	B	C
1			
2			
3		Valores	
4	Rótulos de Linha	Contar de Modelo	Contar de Modelo2
5		1	1
6	Chiconaultia	81	81
7	Deltaforce3	56	56
8	LuxuriousCar	29	29
9	SpaceShuttle	42	42
10	Valentiniana	41	41
11	Total geral	250	250
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			

Figura 6 – Tabela dinâmica de modelo

	A	B	C
1			
2			
3		Valores	
4	Rótulos de Linha	Contar de Modelo	Contar de Modelo2
5		1	1
6	Chiconaultia	81	81
7	Deltaforce3	56	56
8	LuxuriousCar	29	29
9	SpaceShuttle	42	42
10	Valentiniana	41	41
11	Total geral	250	250
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			

Figura 7 – Opções do campo Valor

Selecioneando “Configurações do Campo de Valor...” é possível mudar o formato de apresentação dos dados, com as várias opções mostradas na Figura 8.

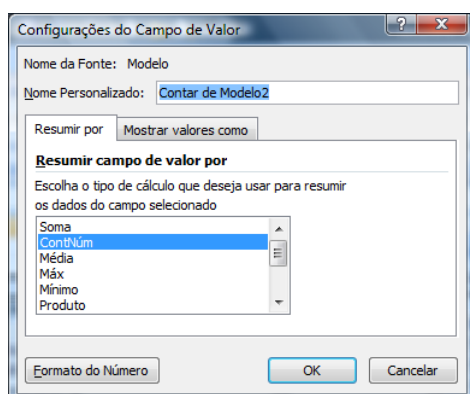


Figura 8 – Configuração do Campo Valor: Resumir por...

Pode-se mudar o tipo de cálculo que se deseja usar para resumir os dados do campo selecionado. Quando a variável é qualitativa a opção padrão é “ContNúm”, contagem. Mas, se a variável for quantitativa, pode-se escolher soma, média, Máx (máximo), Mínimo, Produto e outras funções não visíveis na Figura 8 (desvio padrão amostral e populacional, variância amostral e populacional). É possível ver também a aba “Mostrar valores como”, que uma vez selecionada resulta na Figura 9.

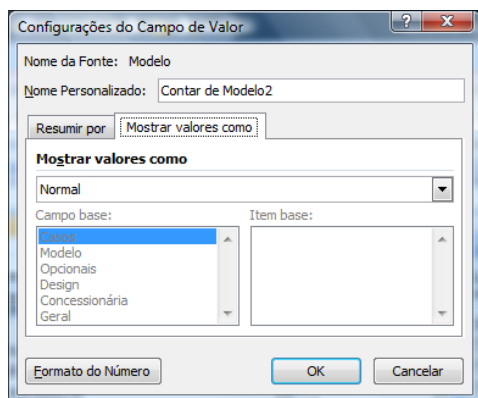


Figura 9 – Configuração do Campo Valor: Mostrar valores como...

A opção padrão é apresentar os valores como “Normal”, ou seja, no caso de contagem para variáveis qualitativas, as frequências absolutas. Mas, pressionando a seta podemos observar as várias opções disponíveis na Figura 10: % de diferença de, soma acumulada, % da linha, % da coluna, % do total entre outras. Como se deseja apresentar as frequências relativas dos valores da variável Tipo, podemos escolher % do total e pressionar OK, a tabela resultante pode ser analisada na Figura 11.

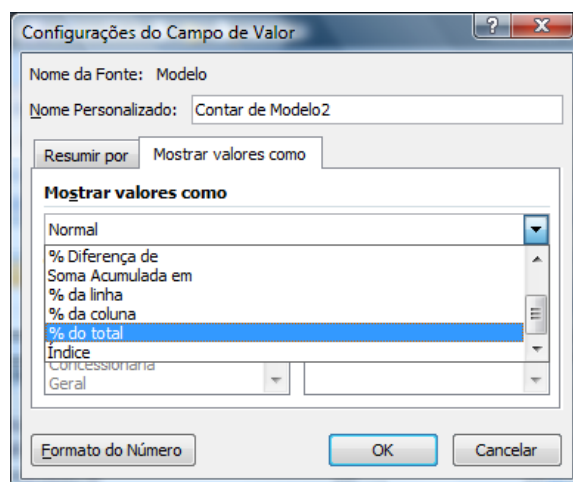


Figura 10 – Configuração do Campo Valor: Mostrar valores como % do total

Rótulos de Linha	Valores	
	Contar de Modelo	Contar de Modelo2
	1	0,40%
Chiconaultla	81	32,40%
Deltaforce3	56	22,40%
LuxuriousCar	29	11,60%
SpaceShuttle	42	16,80%
Valentiniana	41	16,40%
Total geral	250	100,00%

Figura 11 – Tabela dinâmica de Modelo: frequências absolutas e relativas – 1ª versão

As células vazias representam 0,40% do total, perfazendo apenas 1 cliente: é um percentual pequeno (aceita-se até 5%), mas que merece uma reflexão, pois é um dado imprescindível, a empresa precisa descobrir o Modelo deste cliente. Percebe-se a ligeira preponderância dos clientes do Chiconaultla, que representam 32,40% do total (81 clientes). Observa-se também que a tabela apresenta algumas “poluições”, o que pode confundir sua interpretação por leigos: há uma linha para células vazias, “Rótulos de Linha”, “Valores”, “Contar de Modelo”, “Contar de Modelo2”. Todos estes podem ser modificados, os últimos simplesmente por digitação dos nomes de nosso interesse, resultando na Figura 12.

Modelo	Valores	
	Frequências absolutas	%
	1	0,40%
Chiconaultla	81	32,40%
Deltaforce3	56	22,40%
LuxuriousCar	29	11,60%
SpaceShuttle	42	16,80%
Valentiniana	41	16,40%
Total geral	250	100,00%

Figura 12 – Tabela dinâmica de Modelo: frequências absolutas e relativas – 2ª versão

A remoção da linha com as células vazias, porque sua influência é muito reduzida, pode ser feita pressionando a seta ▼ ao lado de Modelo, levando às opções da Figura 13.

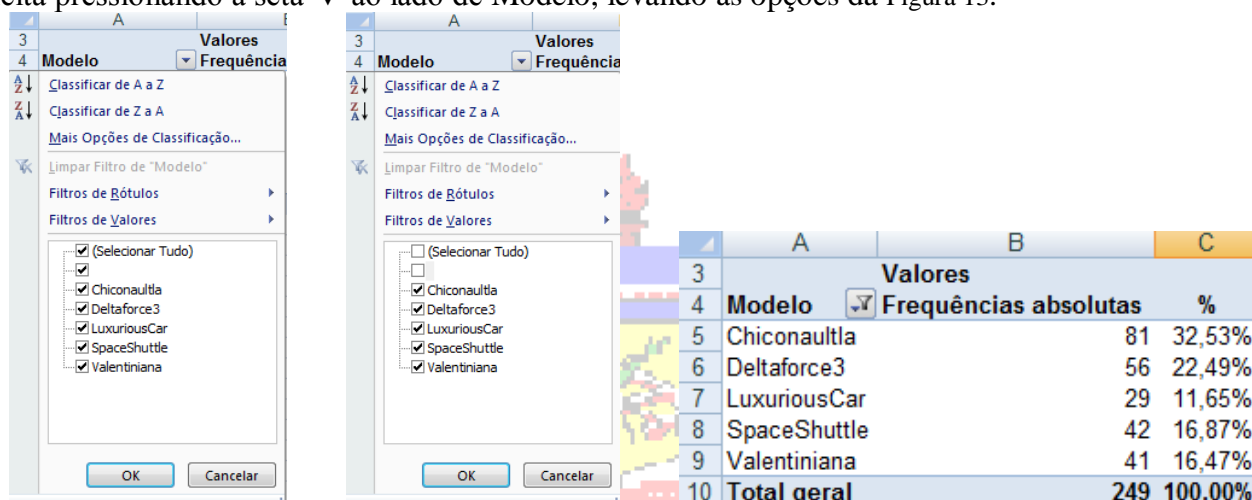


Figura 13 – Opções de exibição da tabela de Modelo: com remoção das células vazias

Basta desmarcar a opção de exibição das células vazias, pressionar OK, e obtém-se a tabela final, com os valores percentuais já recalculados em relação ao novo total (249).

Suponha que há interesse em construir um gráfico para a tabela da variável qualitativa Modelo, Figura 13. As opções possíveis são gráfico de barras (colunas no Excel) e em setores.

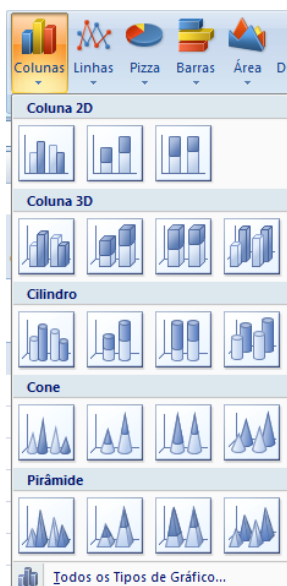
Gráfico de barras (colunas)

Ponha o cursor em algum ponto da tabela dinâmica de Modelo, e procure no menu Inserir por gráficos, como na Figura 14.



Figura 14 – Opções de gráficos

Clique no tipo de gráfico e, em seguida, clique no subtipo de gráfico que deseja usar. No nosso caso, queremos “Colunas” (barras verticais). Se pressionarmos a seta imediatamente abaixo de “Colunas” é possível ver todos os subtipos disponíveis (Figura 15).



Há opção de colunas 2D, 3D, cilindro, cone e pirâmide. Embora os gráficos 3D sejam mais atraentes, eles podem levar a interpretações equivocadas devido aos diferentes ângulos das colunas e eventuais sombras geradas. Assim, recomenda-se o uso de colunas 2D. O primeiro subtipo é o de colunas agrupadas, que utilizam as frequências absolutas como alturas das colunas, para permitir a comparação dos valores. O segundo subtipo é o de colunas empilhadas, e o terceiro colunas empilhadas 100%: estes dois últimos são mais apropriados para o cruzamento de duas variáveis qualitativas, usando frequências absolutas ou percentuais por linha ou coluna, diretamente.

Para a tabela dinâmica de Tipo vamos escolher Colunas 2D, colunas agrupadas, resultando na Figura 16.

Figura 15 - Subtipos de gráficos de colunas

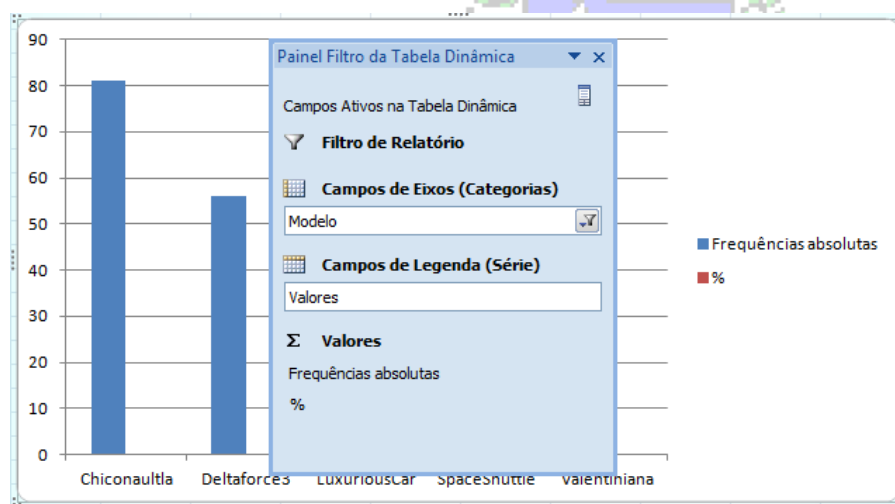


Figura 16 - Gráfico de colunas de Modelo - 1a versão

A ação padrão do Excel 2007 é pôr o gráfico como objeto na planilha onde estão os dados. Observe que aparece no gráfico a legenda “Frequências absolutas” e “%”, mas que estas últimas estão “invisíveis”, pois os percentuais são muito pequenos em relação às frequências absolutas. Há também a caixa “Painel Filtro da Tabela Dinâmica”.

Na parte superior de “Painel Filtro da Tabela Dinâmica” há “Campos Ativos na Tabela Dinâmica”: precisamos pressioná-lo para remover os % do gráfico, com a finalidade de torná-lo mais claro. **Mas, ao fazer isso vamos removê-los da tabela dinâmica original também.** Assim, se não quiser perdê-la, copie a tabela e cole-a *como valores* em algum outro ponto da planilha, e construa o gráfico a partir da cópia. Pressionando “Campos Ativos na Tabela Dinâmica”, chegamos à Figura 17:

- selecionando % no campo Σ Valores, vamos obter a Figura 17;
- escolhendo “Remover campo”, desaparecerá % do gráfico e da tabela dinâmica original de Modelo, resultando na Figura 19.

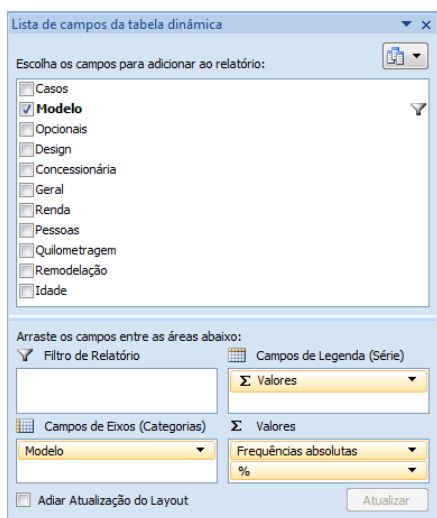


Figura 17 - Campos do gráfico dinâmico

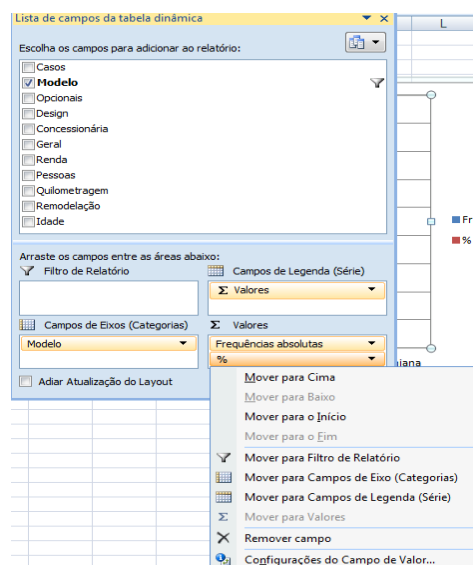


Figura 18 - Opções do campo %

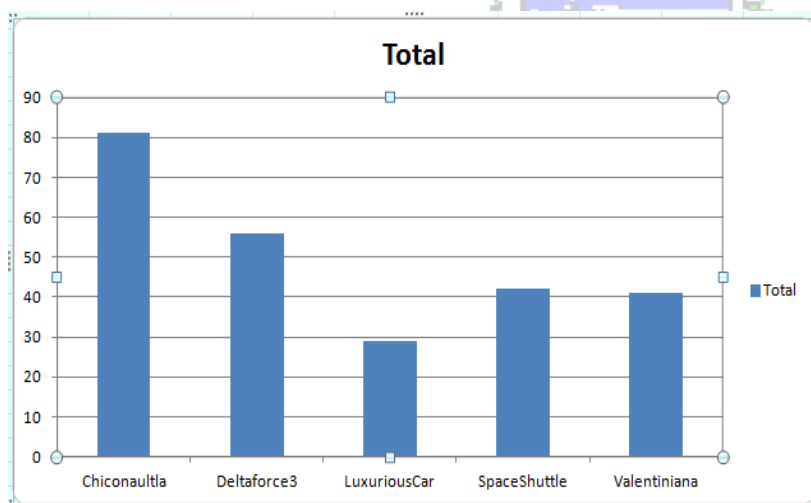


Figura 19 - Gráfico de colunas de Tipo - 2a versão

Observe que o gráfico está com o título genérico “Total”, e apresenta uma legenda desnecessária. Também não há título no eixo vertical, para indicar que se trata de frequências absolutas da variável Modelo.

Para modificar título geral basta clicar sobre ele e reescrevê-lo: Modelo de veículo. Além disso, ao selecionar o gráfico com o mouse surge o conjunto de opções “Ferramentas de Gráfico Dinâmico”.

As opções mais interessantes neste momento são “Design” (Figura 20) e “Layout” (Figura 21).

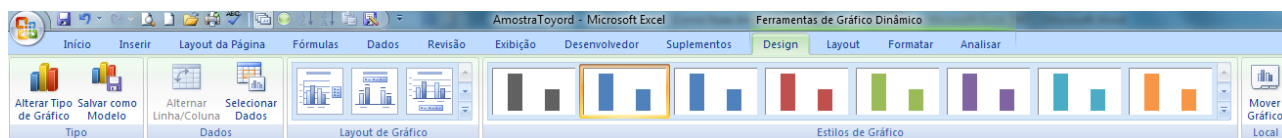


Figura 20 - Opções de design de gráfico dinâmico de colunas

Podemos alterar o tipo de gráfico, selecionar dados, mudar o estilo do gráfico e mesmo movê-lo para outro local, uma planilha específica, por exemplo, para permitir melhor visualização.



Figura 21 - Opções de layout de gráfico dinâmico de colunas

Nas opções de layout podemos modificar formato e posição dos títulos (do gráfico e dos eixos), a área de plotagem, adicionar Linha de Tendência e modificar a Legenda. No nosso caso, não queremos exibi-la, então, ao selecionar Legenda obtemos a Figura 22.

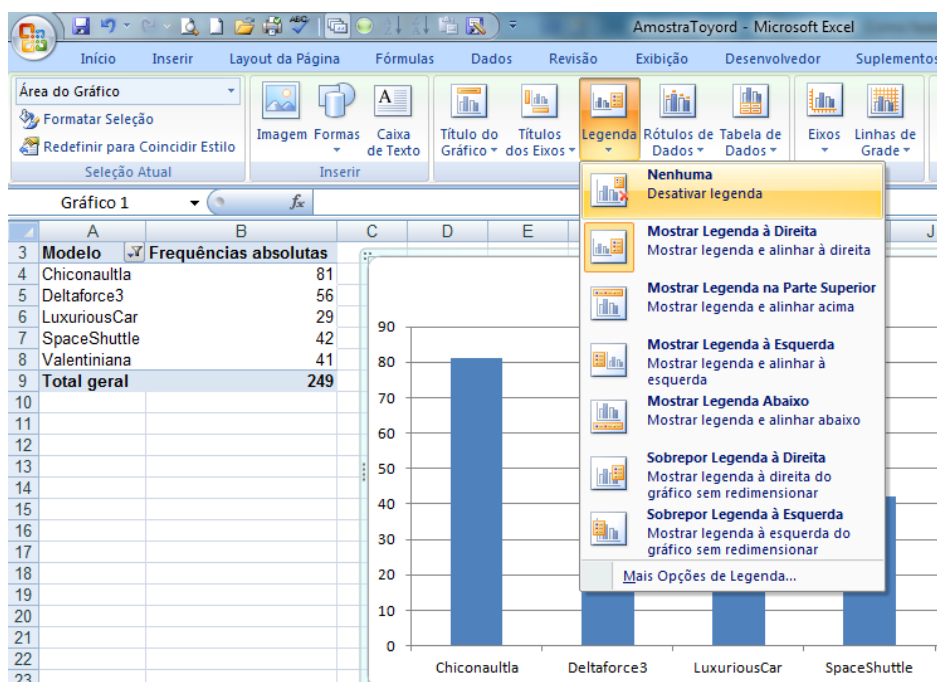


Figura 22 - Opções de legenda do gráfico dinâmico

Escolhendo “Nenhum” na Figura 22 a legenda desaparecerá. Mas, precisamos acrescentar um título ao eixo vertical, selecionando “Título dos Eixos” na Figura 21, e a opção “Título do Eixo Vertical Principal” chegamos à Figura 23.

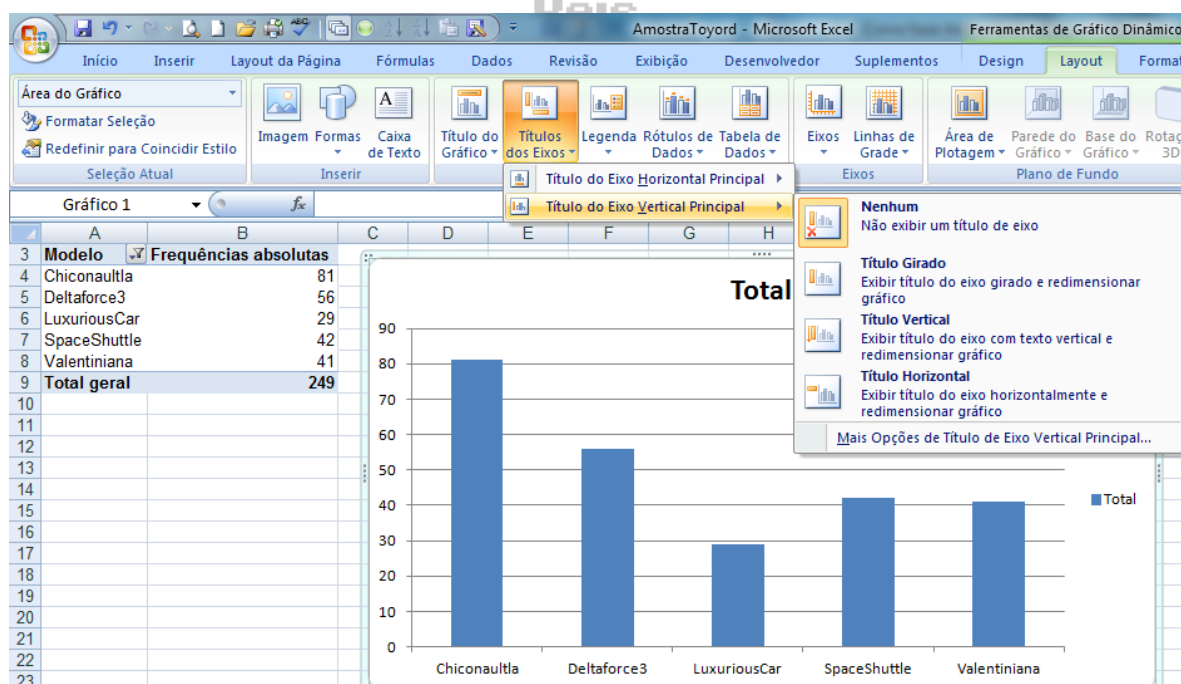


Figura 23 - Opções de Título do Eixo Vertical Principal em Gráfico Dinâmico

“Título Girado” permitirá que o nome acompanhe o eixo, e com esta seleção o Excel 2007 cria automaticamente uma caixa de texto chamada “Título do eixo”, que pode ser modificada

exatamente como o título do gráfico. O resultado final, após a remoção da legenda e acréscimo dos títulos está na Figura 24.

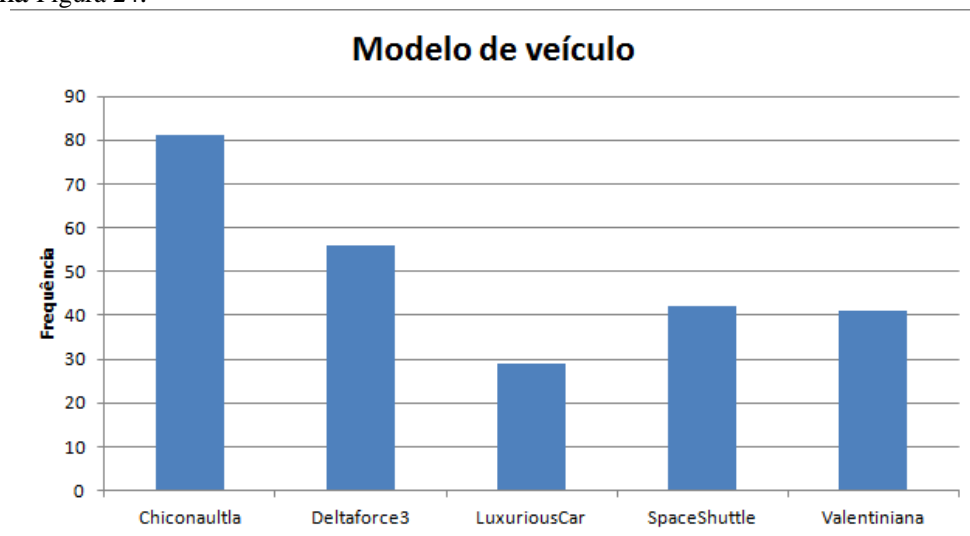


Figura 24 - Gráfico de colunas de Modelo de veículo - Final

A interpretação do gráfico é a mesma da tabela. A principal vantagem do gráfico é a **rapidez na apreensão da informação**: em uma simples olhada o leitor pode compreender o básico da distribuição de frequências, o que muito útil em apresentações ou conferências, onde geralmente há pouco tempo disponível. Mas, a mesma rapidez de apreensão pode nos levar a erros de interpretação se o gráfico for modificado (com ou sem segundas intenções) para causar um determinado efeito. No gráfico da Figura 24 percebe-se uma diferença entre os modelos, com preponderância do Chiconaultla (32,40%) sobre os demais: não se trata de uma grande discrepância e o gráfico mostra isso porque a escala do **eixo vertical começa no zero**. Se não começasse no zero, as posições relativas das alturas das colunas seriam mudadas, o que poderia levar a interpretações equivocadas. Não obstante, pode ser necessário mudar a escala do gráfico. O primeiro passo é selecionar o eixo a mudar e pressionar o botão direito do mouse, levando à Figura 25.

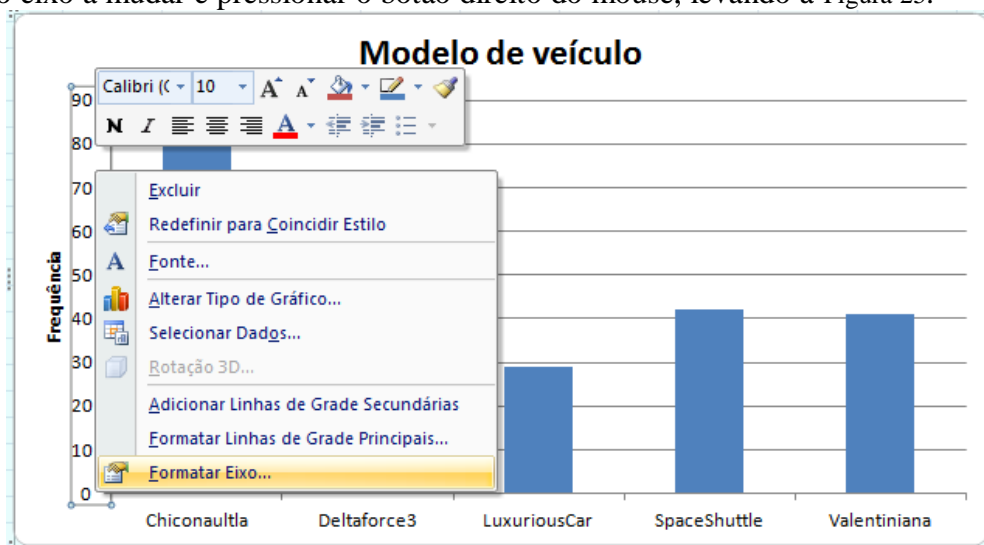


Figura 25 - Opções de formatação de gráfico e eixo

Pressionando “Formatar Eixo” nos leva à Figura 26. Automaticamente o Excel 2007 escolhe a escala mais apropriada de acordo com os dados de origem (Opções de Eixo: Automático). Mas, podemos trocar os limites e as unidades da escala. Na Figura 26 trocamos o mínimo para Fixo, indicando que o eixo vertical começará em 20. O gráfico resultante é mostrado na Figura 28.

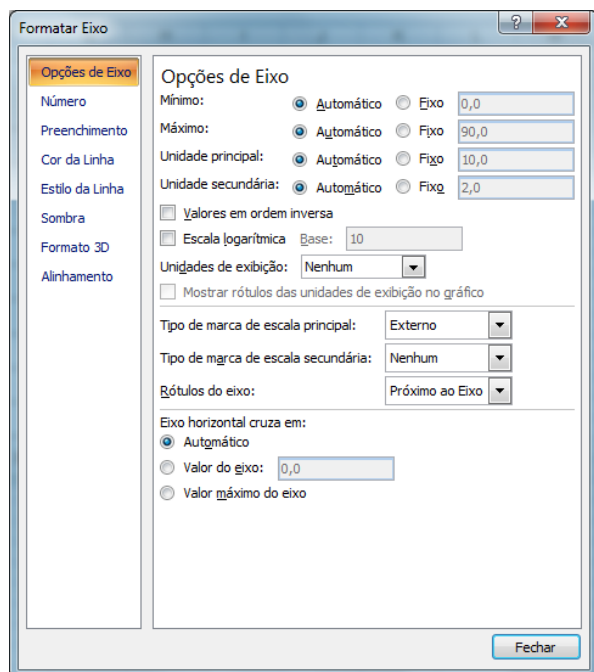


Figura 26 - Opções de Eixo

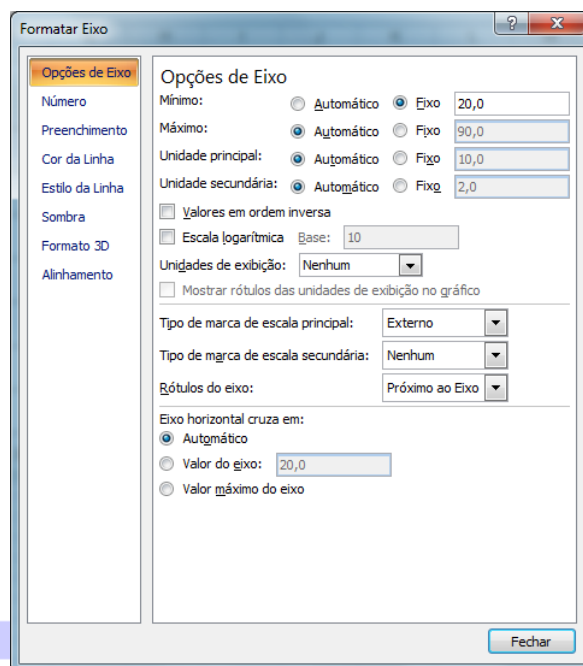


Figura 27 - Mudança de escala no eixo vertical

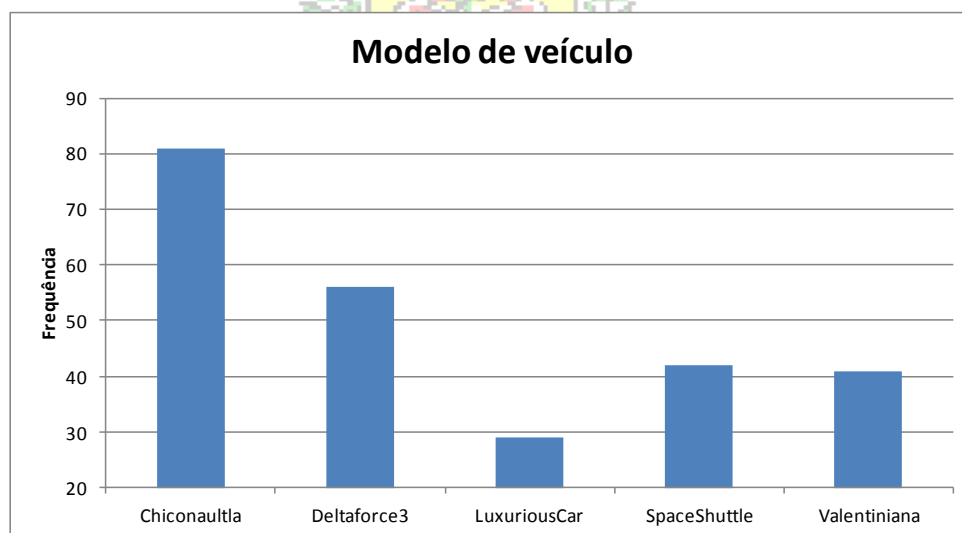


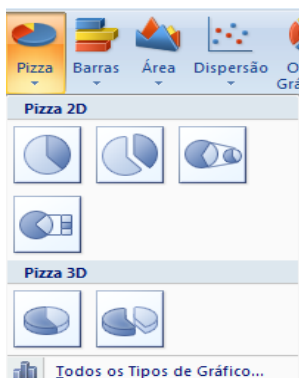
Figura 28 - Gráfico de colunas de Tipo de serviço – escala vertical modificada

Observe que a diferença entre os modelos parece um pouco mais acentuada agora. O leitor desatento pode achá-la maior do que é na realidade, pois seu olhar é “atraído” para as colunas, muito poucos prestarão atenção no valor inicial da escala vertical.

Gráfico de setores

O gráfico de setores pode ser construído da mesma forma que o gráfico em colunas, a partir da tabela dinâmica de uma variável qualitativa. A diferença é que ele apresentará automaticamente os percentuais de cada valor que ela pode assumir em relação ao total geral.

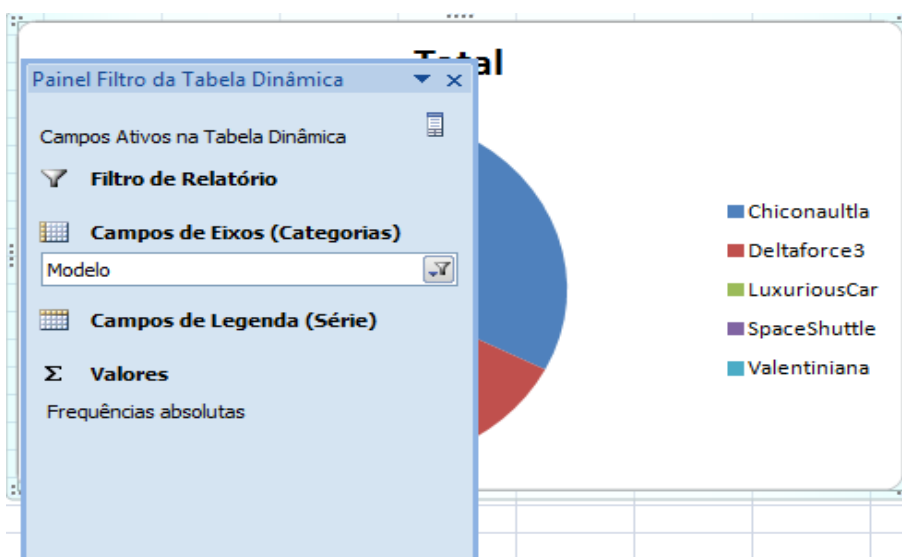
Ponha o cursor em algum ponto da tabela dinâmica de Tipo, e procure no menu Inserir por gráficos, como na Figura 14, mostrada anteriormente. Clique no tipo de gráfico e, em seguida, clique no subtipo de gráfico que deseja usar. Agora, queremos “Pizza”. Se pressionarmos a seta imediatamente abaixo de “Pizza” é possível ver todos os subtipos disponíveis (Figura 29).



Há opção de Pizza 2D e 3D. A mesma recomendação sobre os gráficos 3D feitas nos gráficos de colunas permanecem válidas, além de poderem ser estendidas para os casos de gráficos em pizza com fatias “explodidas” (destacada, no Excel 2007). Assim, recomenda-se o uso de Pizza 2D simples. Um inconveniente do gráfico em pizza no Excel é a impossibilidade de usá-lo quando houver cruzamento de variáveis qualitativas.

Para a tabela dinâmica de Tipo vamos escolher Pizza 2D resultando na Figura 30.

Figura 29 - Subtipos de gráficos em Pizza



Como a tabela dinâmica já não tinha mais o campo %, o gráfico apresenta apenas as frequências absolutas, convertidas em arcos do círculo (o total corresponde a 360° e as frequências absolutas de cada modelo aos respectivos arcos). Precisamos mudar o Título do gráfico (basta clicar sobre “Total” e reescrever, e acrescentar os %, que não estão visíveis).

Figura 30 - Gráfico em pizza de Modelo de veículo - 1a versão

Para acrescentar os percentuais é preciso selecionar o gráfico e observar novamente as “Ferramentas do Gráfico Dinâmico”. As opções de Design na Figura 31(veja que são adaptadas ao gráfico de Pizza), e as opções de Layout na Figura 32, incluindo “Rótulos de Dados”.

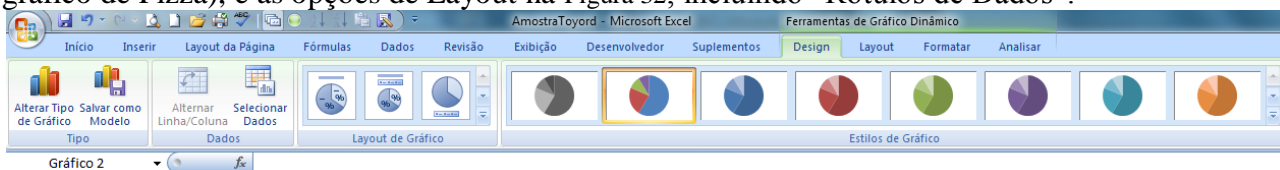


Figura 31 - Opções de design de gráfico dinâmico em Pizza

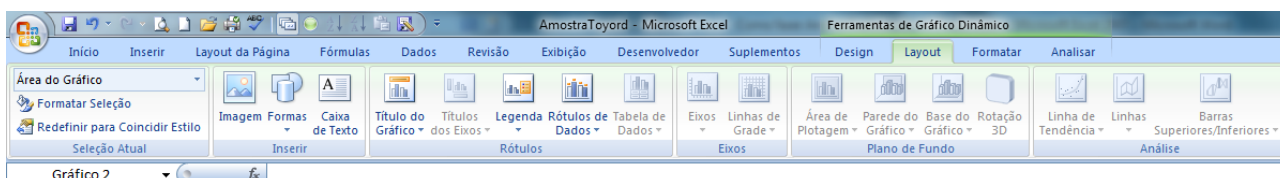


Figura 32 - Opções de layout de gráfico dinâmico em Pizza

Pressionando na seta ao lado de Rótulo de Dados chegamos à Figura 33.

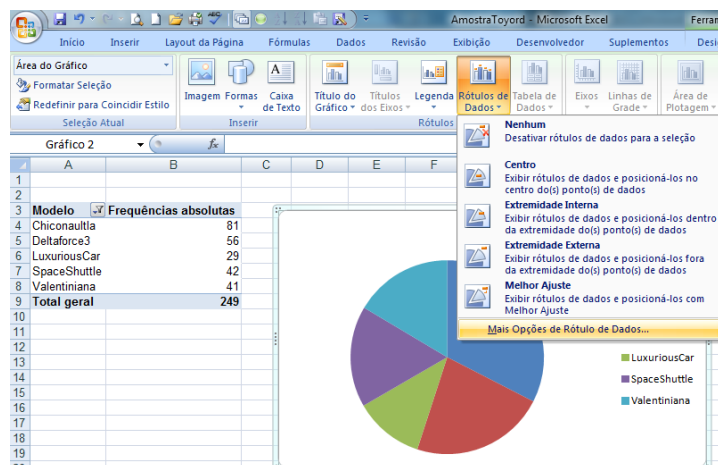


Figura 33 - Opções de Rótulo de Dados para Gráfico em Pizza

As opções se referem inicialmente à posição onde serão colocados os rótulos, destacando-se Centro e Melhor Ajuste. Mas, precisamos de mais opções de Rótulos de Dados para acrescentar os percentuais de interesse. Pressionando “Mais Opções de Rótulos de Dados”, o resultado é a Figura 34. Em “Opções de Rótulo” é possível marcar Porcentagem e escolher sua posição como Melhor Ajuste, finalmente obtendo o gráfico da Figura 35.

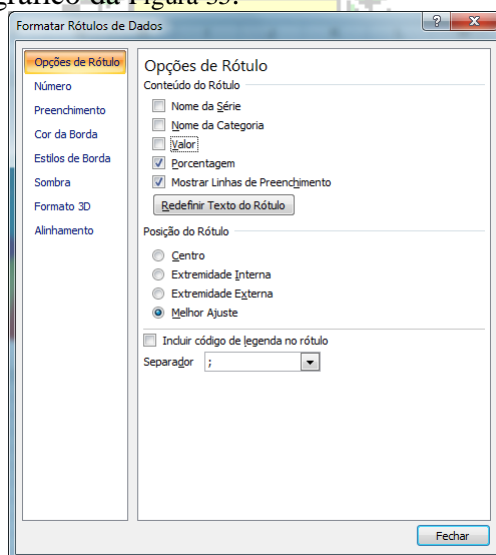


Figura 34 - Mais opções de Rótulo de Dados

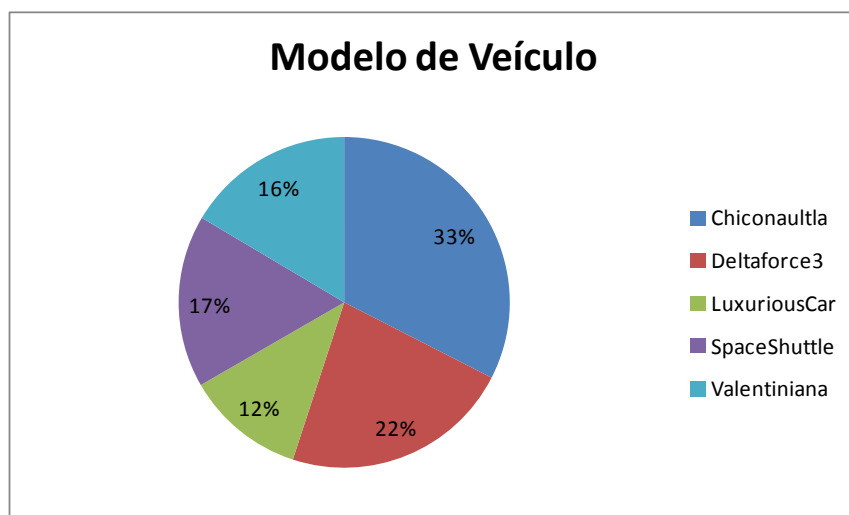


Figura 35 - Gráfico em pizza de Modelo de Veículo - Final

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

É muito comum o estudo de duas variáveis qualitativas. Pode-se representar essa distribuição conjunta de frequências através de uma **tabela de contingências**, para estudar a sua associação. Através de uma tabela dinâmica podemos rapidamente construí-la.

O início do processo é semelhante ao visto na seção anterior: posicione o cursor em algum ponto do conjunto de dados, e procure no menu Inserir por Tabela Dinâmica, como na Figura 1. Mas, ao invés de pôr a tabela em uma nova planilha vamos escolher uma célula da mesma planilha onde está a tabela deModelo, como mostrado na Figura 36.

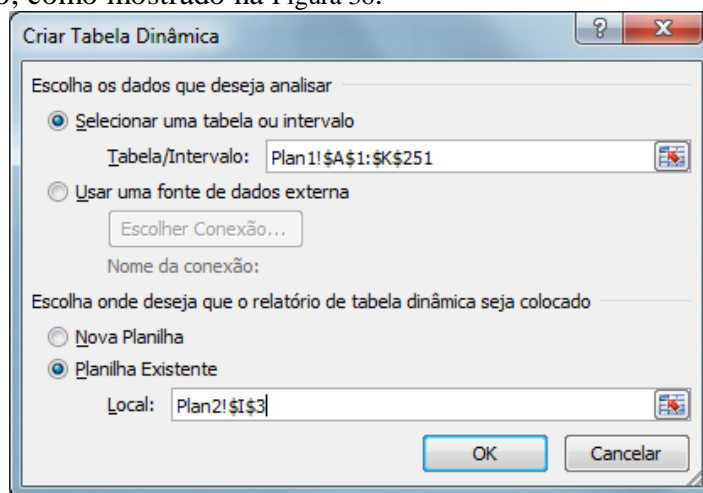


Figura 36 – Célula de destino de uma tabela dinâmica

Pressionando OK, chega-se à uma situação parecida a da Figura 4. Mas, agora vamos colocar duas variáveis na Tabela, Modelo na Linha, e Geral (opinião geral) na Coluna, e apenas uma delas (qualquer uma) em Valores. Como na Figura 37.

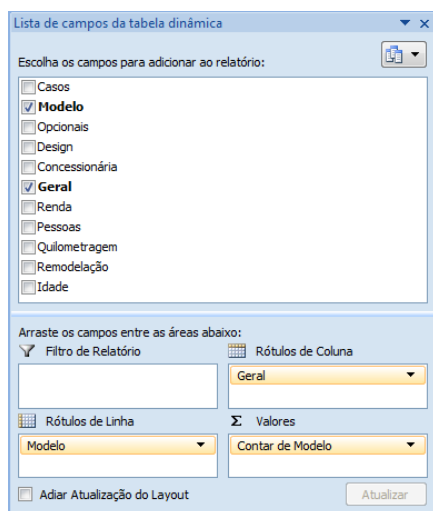


Figura 37 - Tabela dinâmica de Modelo e Geral – Layout inicial

Ao construir uma tabela de contingências usualmente não temos interesse apenas nas frequências de cada cruzamento. Foi visto na seção 1.1 que a tabela dinâmica pode incluir percentuais também. Vamos aproveitar a oportunidade e incluir três tipos de percentuais: em relação aos totais das linhas, em relação aos totais das colunas, e em relação ao total geral dos dados. Basta arrastar a variável Modelo mais três vezes ao campo “Valores”, Figura 38. Posteriormente, podemos modificar as configurações do campo Valor, com as devidas adaptações, tal como da Figura 7 a Figura 10. Ver Figura 40. A tabela final resultante está na Figura 41, já com as modificações apropriadas (incluindo a remoção das células vazias das duas variáveis).

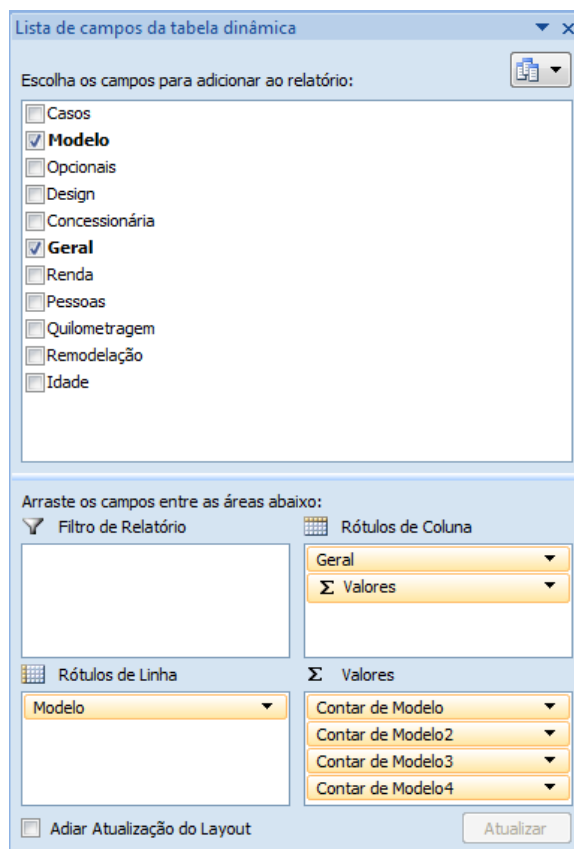


Figura 38 - Tabela dinâmica de Modelo e Geral – Layout com frequências absolutas e todos os %.

Da forma como está o layout da Figura 38 a tabela não ficará com um bom aspecto, dificultando a sua interpretação. Precisamos mover o botão Σ Valores do campo Rótulos de Coluna (ver Figura 38) para o campo Rótulos de Linha, resultando na Figura 39.

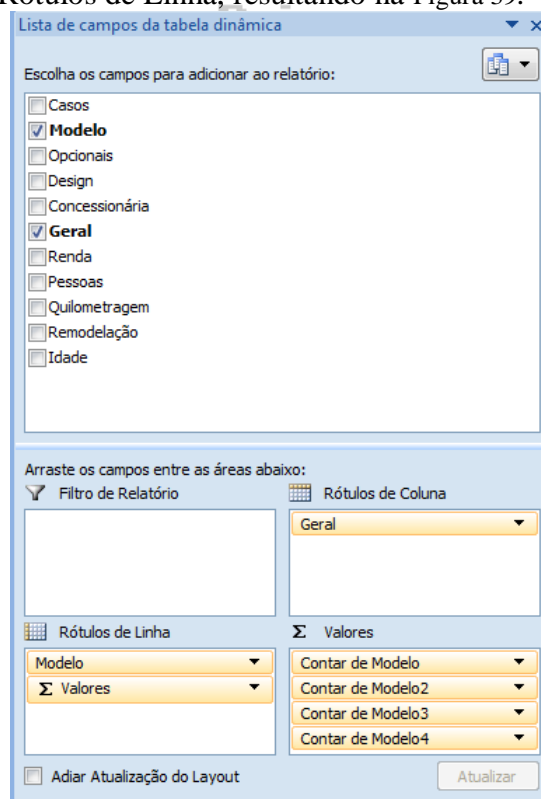


Figura 39 - Tabela dinâmica de Tipo x Opinião – Layout com frequências absolutas e todos os %, modificado

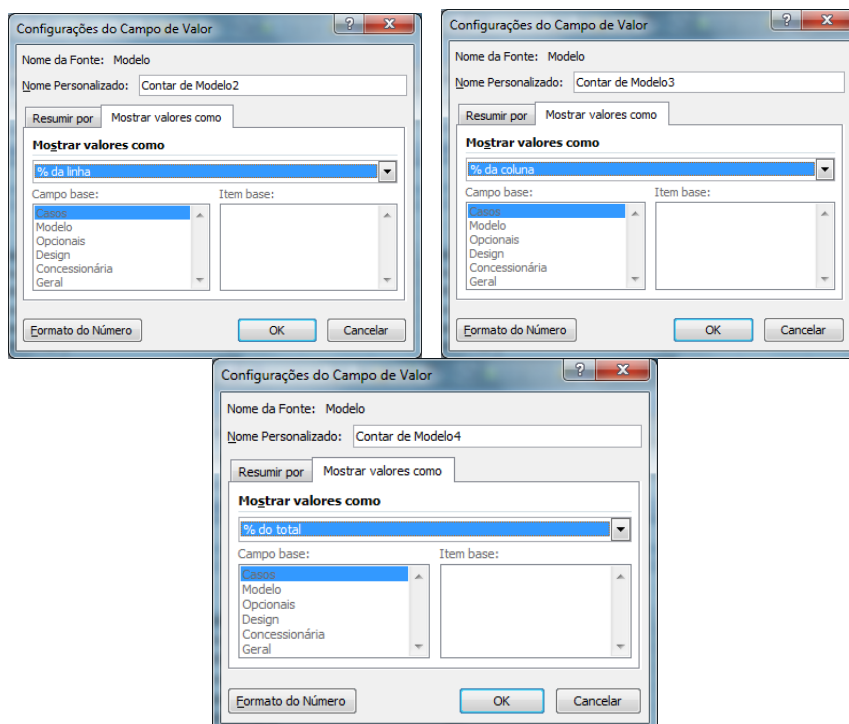


Figura 40 - % por linha, coluna e total

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

Figura 41 - Tabela dinâmica de Modelo x Geral (opinião geral) (sem células vazias)

Antes de analisar os resultados vamos fazer uma pequena modificação na tabela. Observe que o Excel ordena os valores das variáveis de maneira alfabética: Bastante satisfeito, Insatisfeito, Muito insatisfeito e Satisfeito. Repare também que o ideal seria começar da “melhor” para a “pior”

opinião (ou vice-versa). Para conseguir isso basta mudar a ordem dos valores. Vamos fazer isso para a variável Geral, mudando a posição de Satisfeito para a esquerda: basta selecionar Satisfeito, e pressionar o botão direito do mouse, resultando na Figura 42.

Modelo	Geral	Bastante satisfeito	Insatisfeito	Muito insatisfeito	Satisfeito	Total geral
Chiconaultla						
Frequência			11	69		
% por linha		0,00%	13,58%	85,19%		
% por coluna		0,00%	16,67%	62,73%		
% do total		0,00%	4,42%	27,71%		
Deltaforce3						
Frequência			22	29		
% por linha		0,00%	39,29%	51,79%		
% por coluna		0,00%	33,33%	26,36%		
% do total		0,00%	8,84%	11,65%		
LuxuriousCar						
Frequência		19	1			
% por linha		65,52%	3,45%	0,00%		
% por coluna		59,38%	1,52%	0,00%		
% do total		7,63%	0,40%	0,00%		
Space Shuttle						
Frequência		10	14	1		
% por linha		23,81%	33,33%	2,38%		
% por coluna		31,25%	21,21%	0,91%		
% do total		4,02%	5,62%	0,40%	6,83%	16,87%
Valentiniana						
Frequência		3	18	11	9	41
% por linha		7,32%	43,90%	26,83%	21,95%	100,00%
% por coluna		9,38%	27,27%	10,00%	21,95%	16,47%
% do total		1,20%	7,23%	4,42%	3,61%	16,47%
Total Frequência		32	66	110	41	249
Total % por linha		12,85%	26,51%	44,18%	16,47%	100,00%

Figura 42 - Tabela dinâmica Modelo x Geral - opções de formatação

Observe a opção Mover na Figura 43, selecionando-a vamos à Figura 44.

Modelo	Geral	Bastante satisfeito	Insatisfeito	Muito insatisfeito	Sa
Chiconaultla					
Frequência		11	69		
% por linha		0,00%	13,58%	85,19%	
% por coluna		0,00%	16,67%	62,73%	
% do total		0,00%	4,42%	27,71%	
Deltaforce3					
Frequência		22	29		
% por linha		0,00%	39,29%	51,79%	
% por coluna		0,00%	33,33%	26,36%	
% do total		0,00%	8,84%	11,65%	
LuxuriousCar					
Frequência		19	1		
% por linha		65,52%	3,45%	0,00%	
% por coluna		59,38%	1,52%	0,00%	
% do total		7,63%	0,40%	0,00%	
Space Shuttle					
Frequência		10	14	1	
% por linha		23,81%	33,33%	2,38%	
% por coluna		31,25%	21,21%	0,91%	
% do total		4,02%	5,62%	0,40%	6,83%
Valentiniana					
Frequência		3	18	11	9
% por linha		7,32%	43,90%	26,83%	21,95%
% por coluna		9,38%	27,27%	10,00%	21,95%
% do total		1,20%	7,23%	4,42%	3,61%
Total Frequência		32	66	110	41
Total % por linha		12,85%	26,51%	44,18%	16,47%
Total % por coluna		100,00%	100,00%	100,00%	100,00%
Total % do total		12,85%	26,51%	44,18%	16,47%

Figura 43 - Tabela dinâmica Modelo x Geral - mudança de ordem de valores

Há três opções possíveis: mover Geral para Linhas (o que não queremos), mover “Satisfeito” para o início ou mover para a esquerda. Como precisamos mudar apenas Satisfeito escolhemos mover para a esquerda, até estar exatamente após Bastante satisfeito, veja a Figura 44.

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

Figura 44 - Tabela dinâmica de Modelo x Geral

Podemos comparar as opiniões gerais em função do modelo de veículo. Sempre devemos comparar os percentuais dos cruzamentos (por LINHA ou por COLUNA) com percentuais em relação aos totais de Modelo ou Geral.

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 Figura 44 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 como cada opinião se distribui pelos modelos, devemos usar os percentuais por coluna:

- do total da opinião “Bastante satisfeito” *nenhum* percentual foi identificado nos modelos Chiconaultla e DeltaForce3, (não obstante eles representem 32,53% e 22,49% da amostra, respectivamente) enquanto que 59,38% das manifestações estão concentradas no modelo LuxuriousCar e 31,25% no SpaceShuttle (que, por sua vez, representam apenas 11,65% e 16,87%

da amostra, respectivamente). Isso significa que, embora menos vendidos, estes modelos estão agradando seus compradores.

- do total da opinião “Muito insatisfeito”, nada menos do que 62,73% são oriundas do modelo Chiconaultla, contra 0,91% de SpaceShuttle e *nenhum* percentual de LuxuriousCar.

Corroborar-se a conclusão da análise dos percentuais por linha, pois os percentuais por opinião (Geral) são significativamente diferentes dos percentuais totais de modelo.

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.

- os clientes bastante satisfeitos com o LuxuriousCar representam 7,63% das respostas totais.

Da mesma forma que para uma variável qualitativa podemos construir um gráfico de barras, mas aqui múltiplas, para representar a tabela da Figura 44. Usando procedimentos análogos aos da Figura 14 à Figura 17 vamos ter:

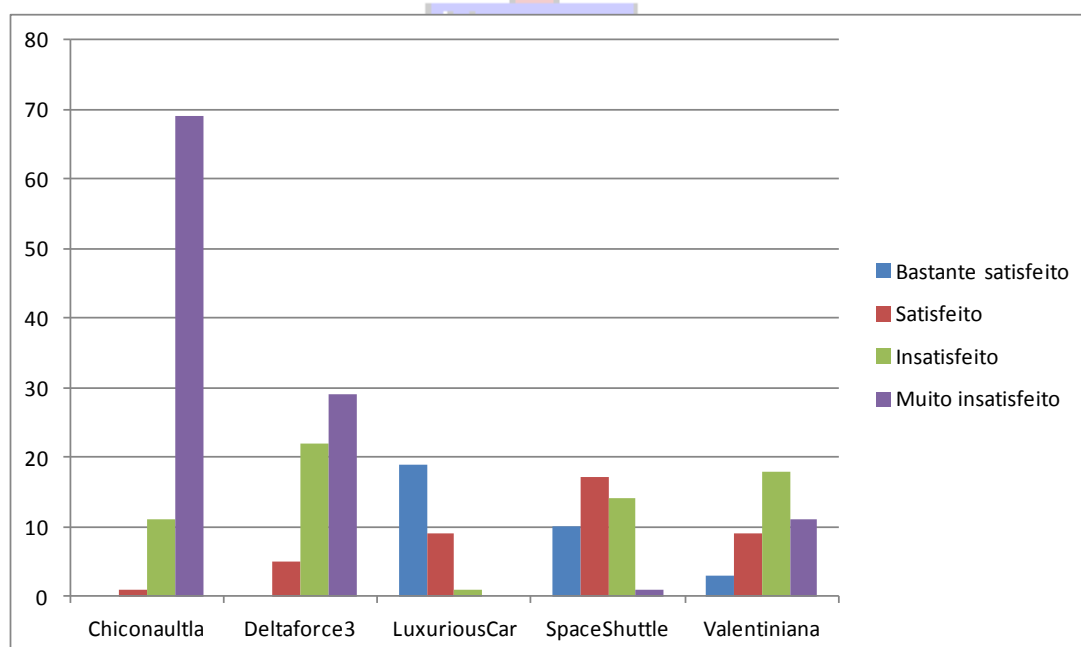


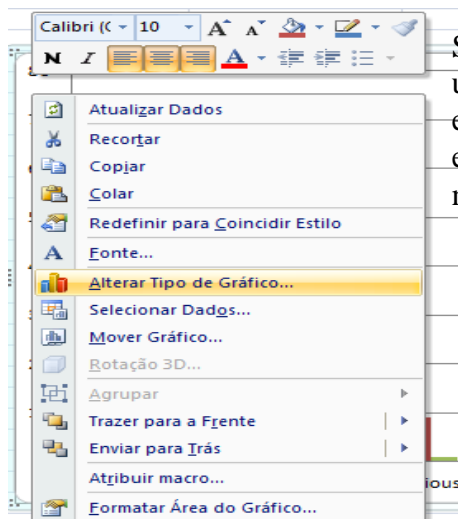
Figura 45 - Gráfico em barras múltiplas da Opinião Geral por Modelo

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 45 e pressionar o botão direito do mouse, o que resulta nas opções mostradas na Figura 46.



Se pressionarmos “Tipo de gráfico” na Figura 46, chegaremos a uma situação semelhante a da Figura 15. Podemos procurar, entre os tipos de gráficos de colunas, o de colunas 100% empilhadas, como mostrado na Figura 47. E pressionando “OK” na Figura 47 resulta o gráfico mostrado na Figura 48.

Figura 46 – Opções de gráfico dinâmico

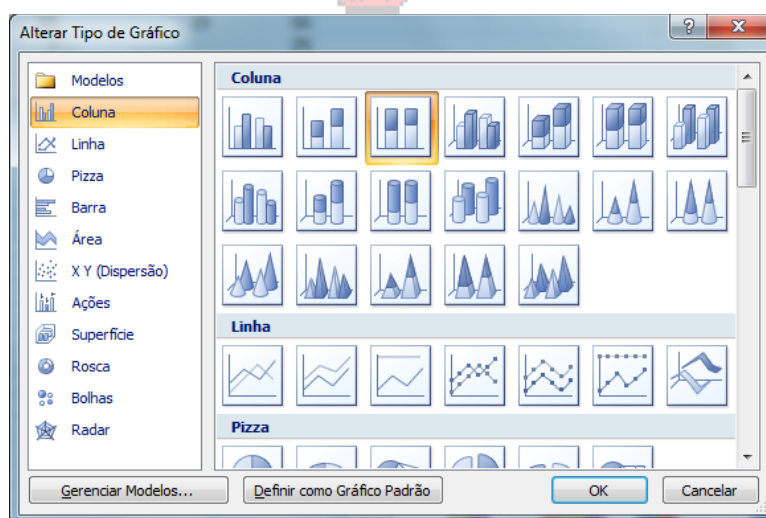


Figura 47 – Opções de gráfico dinâmico – colunas 100% empilhadas

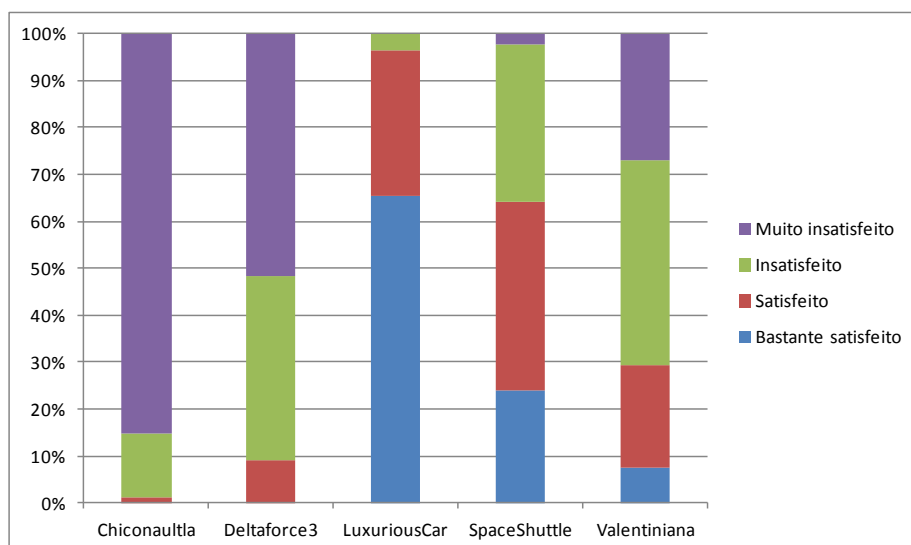


Figura 48 – Gráfico de colunas 100% empilhadas: Opinião geral por Modelo

É possível ver claramente que os percentuais de insatisfeitos e satisfeitos mudam substancialmente de um modelo para outro, exatamente como mostrado na **Erro! Fonte de referência não encontrada.** (percentuais por linha).

2. Procedimentos para variáveis quantitativas

Para variáveis quantitativas há uma maior variedade de procedimentos disponíveis. É preciso distinguir os procedimentos relativos à 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 Tipo de Serviço). 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:

Valores		
Remodelação	Frequência	%
0	2	0,80%
1	57	22,80%
2	123	49,20%
3	59	23,60%
4	9	3,60%
Total geral	250	100,00%

Figura 49 - 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 50.

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

Figura 50 - 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 49.

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.

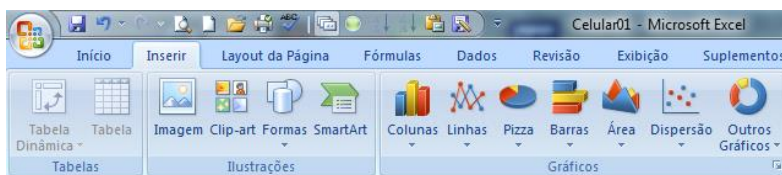


Figura 51 – Opções de gráficos

Nas opções de gráficos não há “Histograma”, mas podemos escolher o gráfico em colunas, pois um histograma não deixa de ser um gráfico de colunas justapostas. Se pusermos o cursor em qualquer ponto da tabela da Figura 49 e selecionarmos Inserir, gráfico, colunas 2D, o resultado será a Figura 52.

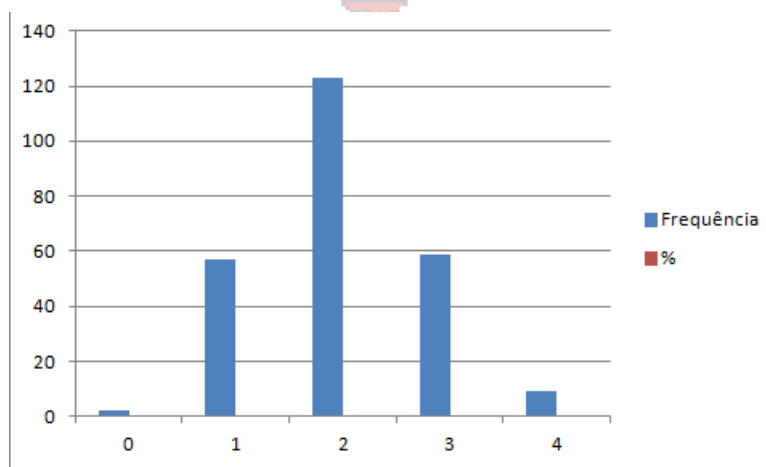


Figura 52 - Histograma de Anos de remodelação - 1a versão

Temos uma situação semelhante a da Figura 16. Podemos aplicar os mesmos procedimentos vistos da Figura 17 à Figura 17, removendo os percentuais e os campos dinâmicos, resultando no gráfico final da Figura 53:



Figura 53 - Gráfico de anos de remodelação - 2a versão

Foi mencionado que as colunas do histograma precisam ser justapostas. Portanto,

precisamos reduzir o espaçamento entre elas a zero. Basta selecionar as colunas e pressionar o botão direito do mouse (Figura 54).



Figura 54 - Opções de formatação de séries de dados - histograma anos de remodelação

Escolhendo "Formatar Série de Dados" surge a Figura 55.

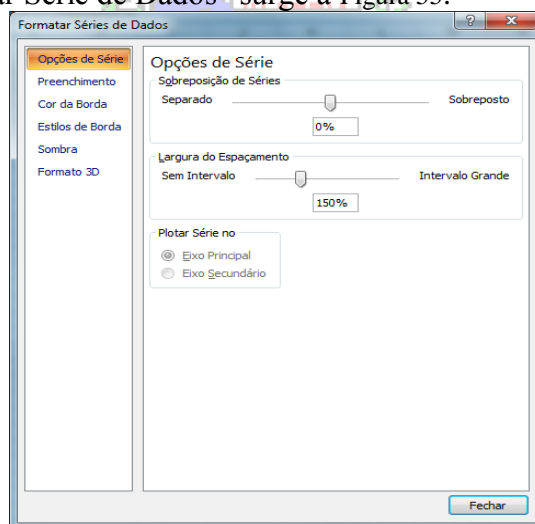


Figura 55 - Opções de formatar séries de dados

Reduzindo o campo "Largura do espaçamento" a Sem intervalo, o histograma correto é mostrado na Figura 118.

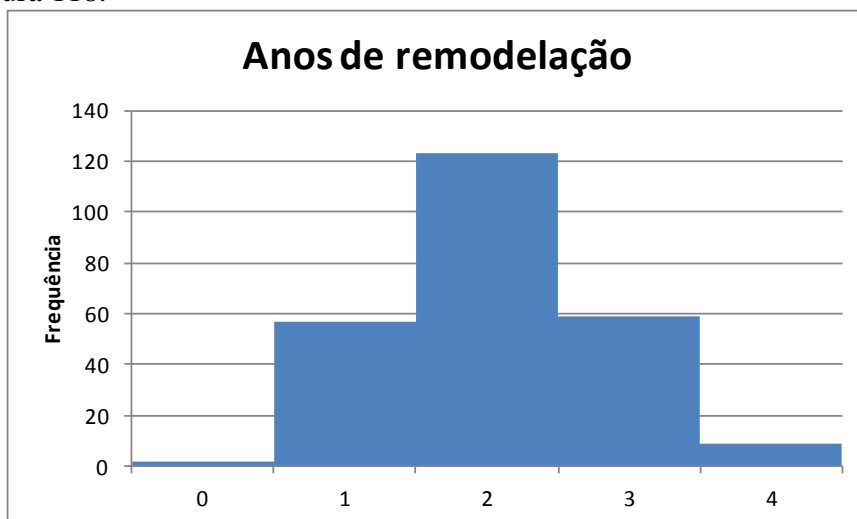


Figura 56 - Histograma de anos de remodelação.

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 necessários mostrados na Unidade 3:

1) Determinar o intervalo do conjunto

Para tanto, temos que identificar os extremos do conjunto, seus valores máximo e mínimo. Vamos colocar o mínimo na célula O2 e o máximo na célula O3, na mesma planilha dos dados. Selecione esta célula com o cursor. Podemos inserir as funções ou escrevê-las diretamente nas células:

	N	O	P		N	O	P
1	Renda			1	Renda		
2	Mínimo	=MÍNIMO(G2:G251)		2	Mínimo	1,795	
3	Máximo			3	Máximo	=MÁXIMO(G2:G251)	

Figura 57 - Mínimo e máximo para a variável Renda

Para calcular o intervalo basta colocar a fórmula respectiva (máximo – mínimo) em uma célula, O4, por exemplo. Veja a Figura 58.

	N	O
1	Renda	
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	=O3-O2

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.

Figura 58 - Intervalo para Renda

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 colocados nas células O5 e O6, respectivamente, como mostrado na Figura 59. 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	Renda	
2	Mínimo	1,795
3	Máximo	86,015
4	Intervalo	84,22
5	Classes	5
6	Amplitude	=O4/5

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 60.

Figura 59 - Amplitude de classes de Renda

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

O limite inferior da primeira classe será o conteúdo da célula N10 (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 61.

	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 61 – 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 62.

	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

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 35,486 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 **Erro! Fonte de referência não encontrada..**

Figura 62 - Classes de Renda

	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 63 – 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 64). 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	V
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)						

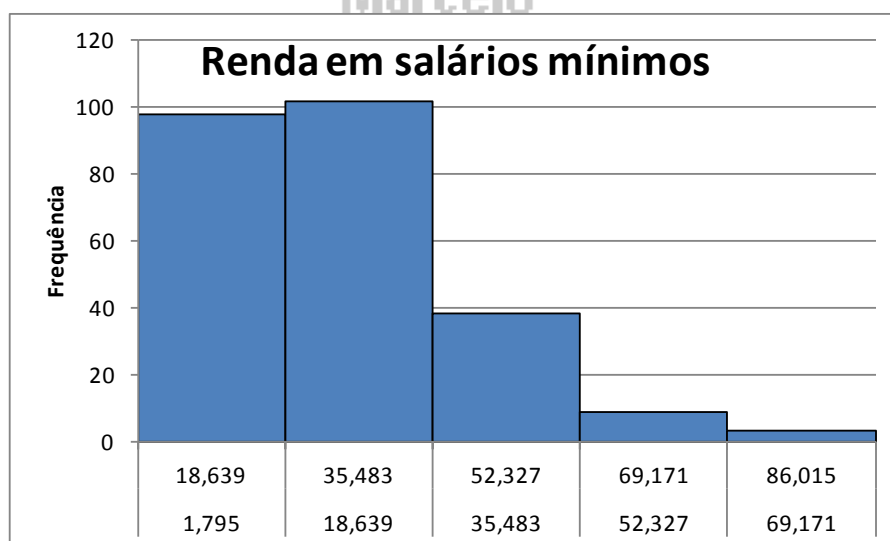
Figura 64 - Obtenção da frequência da última classe de Renda

Observe que na obtenção da frequência 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. Precisamos também definir os pontos médios das classes, que passarão a representá-las, como se as frequências obtidas fossem deles, ver Figura 65.

	N	O	P	Q		N	O	P	Q
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	Frequências	Pontos médios	9	Limite inferior	Limite superior	Frequências	Pontos médios
10	1,795	18,639	98	$=(N10+O10)/2$	10	1,795	18,639	98	10,217
11	18,639	35,483	102		11	18,639	35,483	102	27,061
12	35,483	52,327	38		12	35,483	52,327	38	43,905
13	52,327	69,171	9		13	52,327	69,171	9	60,749
14	69,171	86,015	3		14	69,171	86,015	3	77,593

Figura 65 – 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 66.

**Figura 66 - 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.

A função SE pode ser inserida através do menu Fórmulas (procurando pela opção “Lógica”) ou 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 67.

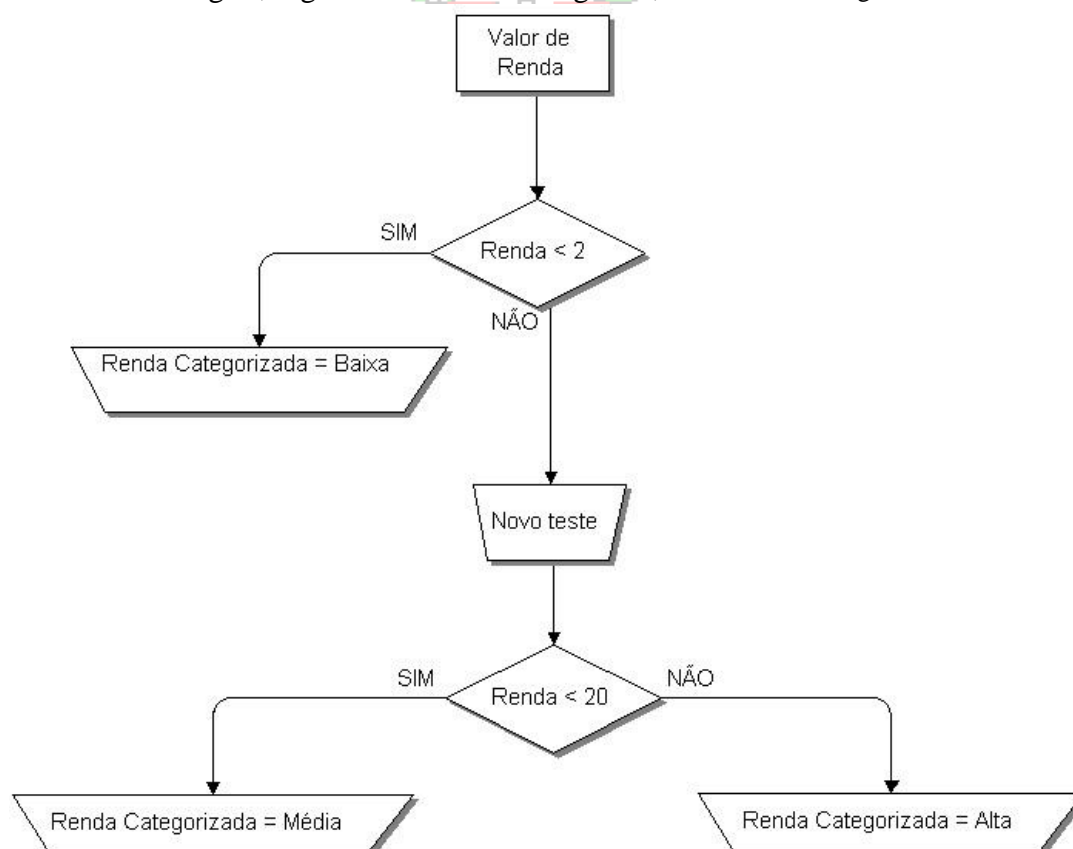


Figura 67 - Procedimento para categorizar a variável Renda
Fonte: elaborada pelo autor.

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 67 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
1	Renda categorizada			
2	=SE(G2<2;"Baixa";SE(G2>20;"Alta";"Média"))			

Figura 68 - 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 parêntese 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 69.

	V	W
1	Renda categorizada	
2	Alta	
3	Alta	
4	Alta	
5	Média	
6	Média	
7	Alta	
8	Alta	
9	Alta	
10	Alta	
11	Alta	
12	Alta	

Figura 69 - Variável Renda Categorizada

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

Rótulos de Linha	Contar de Renda categorizada
Alta	144
Baixa	1
Média	105
Total geral	250

Figura 70 - 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 – Medidas de síntese para uma variável quantitativa

Imagine que queremos fazer uma descrição da variável Renda, calculando as seguintes medidas: média, mediana, moda, desvio padrão, quartil inferior, quartil superior, mínimo e máximo. Todas estas medidas têm funções pré-programadas no Excel, e podem ser inseridas de várias formas. Por exemplo, quando conhecemos a sintaxe da função podemos digitá-la diretamente, como fizemos com máximo e mínimo de Renda (Figura 57). Ou, podemos usar os menus do Excel 2007: basta selecionar “Fórmulas” e surge a Figura 71.

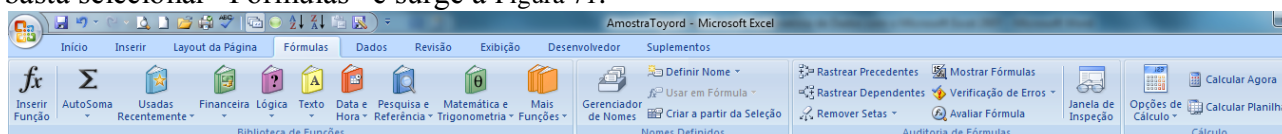


Figura 71 - Fórmulas no Excel 2007

Na extrema esquerda da Figura 71 está a opção “Inserir Função”¹. Observe também as opções: “Usadas Recentemente”, “Financeira”, etc. E, ao final, “Mais Funções”. Se pressionarmos esta última chegamos à Figura 72.

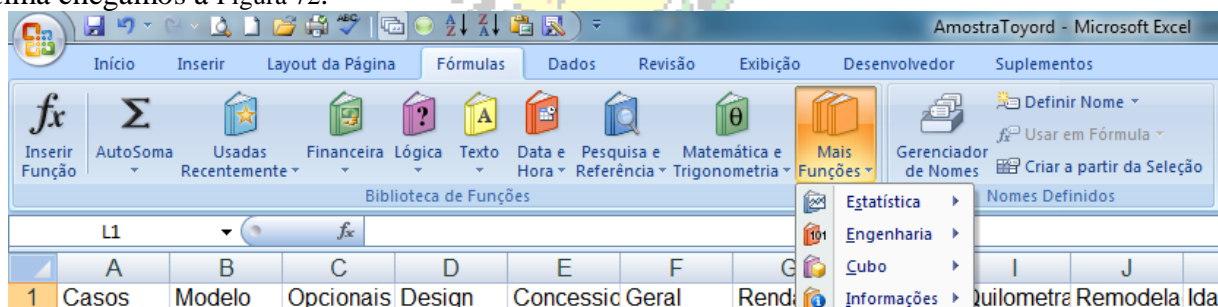


Figura 72 - Fórmulas no Excel 2007 – Mais funções

Dentre as opções vemos “Estatística”, que inclui as medidas de síntese que procuramos. Podemos inserir as medidas usando Inserir Função ou pelo caminho da Figura 72. Mas, antes de tudo, devemos definir as células que armazenarão os resultados. Por conveniência os resultados das medidas serão postos nas células N2 a N8, com os nomes das medidas nas células M2 a M8, tal como mostrado na Figura 73.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Casos	Modelo	Opcionais	Design	Concessão	Geral	Renda	Pessoas	Quilômetros	Remodela	Idade		Medidas	
2	4002	Deltaforce Ar_e_direi	Atualizado	Não causc	Muito insai		24,98	5	415	2	35		Média	
3	3143	SpaceShu AD_Trio_f	Atualizado	Satisfatóri	Satisfeito		24,98	5	597	2	34		Mediana	
4	2658	Valentiniar Ar_e_direi	Ultrapassa	Não causc	Muito insai		23,685	4	594	2	39		Moda	
5	4440	Chiconault AD_Trio_f	Atualizado	Insatisfatóri	Muito insai		19,72	4	422	2	36		Desvio padrão	
6	2018	Deltaforce Ar_e_direi	Atualizado	Não causc	Insatisfeito		12,96	3	503	2	32		Quartil inferior	
7	294	Valentiniar Inexistente	Atualizado	Satisfatóri	Muito insai		40,05	6	604	2	44		Quartil superior	
8	248	Valentiniar AD_Trio_f	Atualizado	Bastante s	Insatisfeito		28,34	5	394	3	28		Mínimo	
9	1527	Valentiniar Ar_e_direi	Atualizado	Muito insai	Bastante s		20,6	4	518	1	45		Máximo	

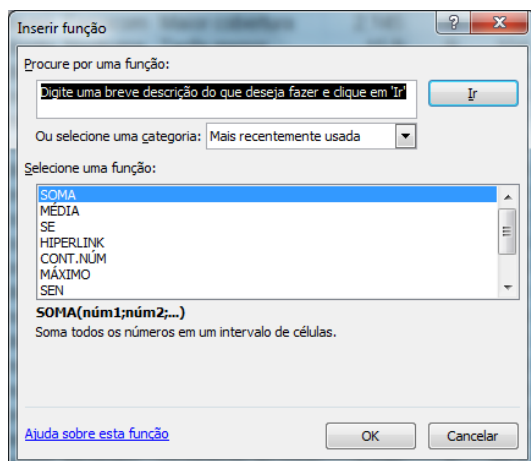
Figura 73 - Medidas de síntese de Renda - 1a parte

¹ Exatamente como nas versões anteriores do Excel.



Medidas de síntese pelo botão Inserir Função

Vamos começar com a média. Posicione o cursor na célula P2. Ao pressionar o botão “Inserir Função” surge a tela da Figura 74.



Se soubermos algo da função podemos digitar parte do seu nome no campo “Procure por uma função” e o Excel apresentará uma lista de candidatas. Ou podemos selecionar uma categoria: a padrão é “Mais recentemente usada”. Se clicarmos sobre a seta ▼ ao lado, vamos ver várias categorias (Figura 75), entre elas a Estatística. Ao escolhermos esta, todas as funções existentes desta categoria serão apresentadas, bastando percorrer a barra de rolagem para encontrar as que procuramos, como, por exemplo, a média.

Figura 74 - Funções disponíveis no Excel

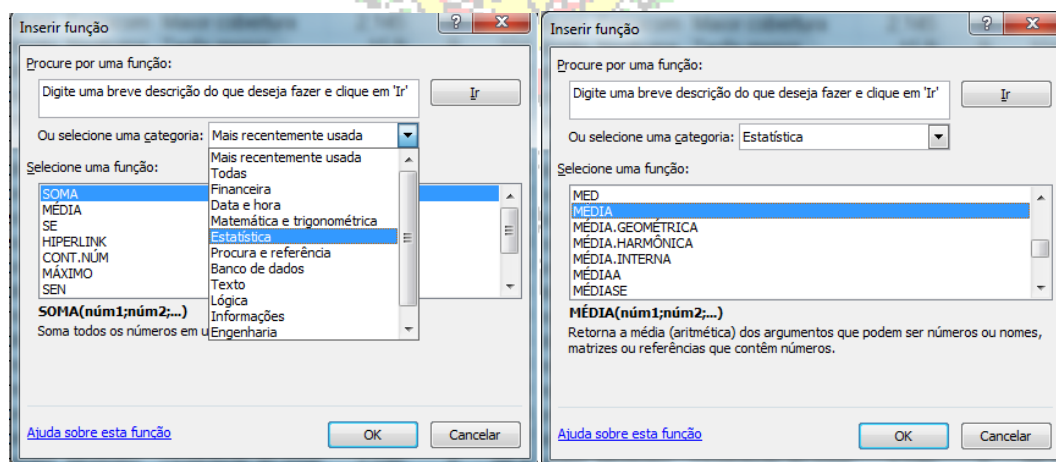


Figura 75 - Funções estatísticas no Excel - 1a abordagem

Para cada função o Excel apresenta a sua sintaxe (os argumentos necessários) e uma explicação sobre o que ela faz. Se pressionarmos OK, passamos à tela da Figura 76.

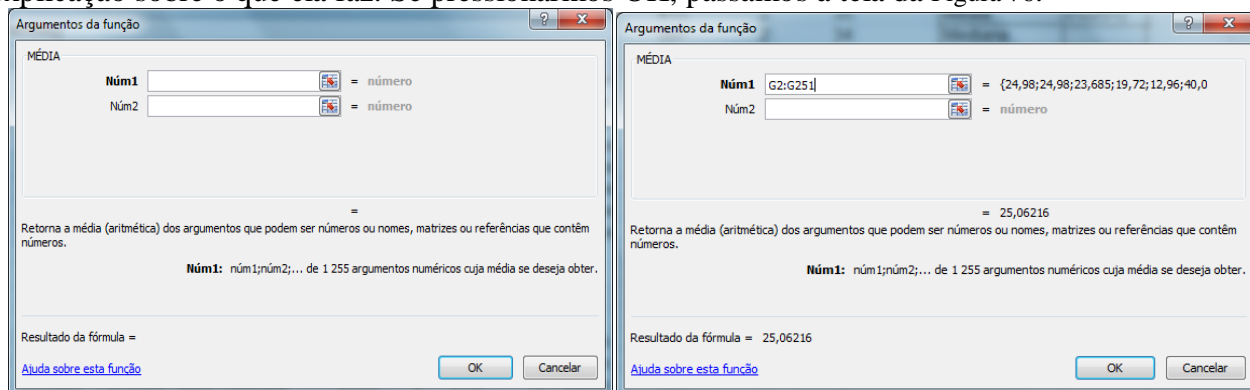


Figura 76 - Argumentos da função Média (para Renda) - 1a versão

Os dados de Renda estão nas células G2 a G251². Então, ou selecionamos os dados diretamente (pressionando a seta vermelha ao lado dos campos “Núm1” ou “Núm2” e escolhendo as células com o mouse) ou escrevemos o intervalo G2:G251. Os dois pontos indicam para o Excel que todas as células do intervalo precisam ser consideradas no cálculo, mas a função desconsidera células vazias. Antes mesmo de pressionar OK o Excel já apresenta o resultado: 25,0622 salários mínimos (cerca de R\$ 15600, em janeiro de 2012). Se repetirmos o processo para as outras medidas, vamos completar os cálculos, os procedimentos estão mostrados da Figura 77 à Figura 80.

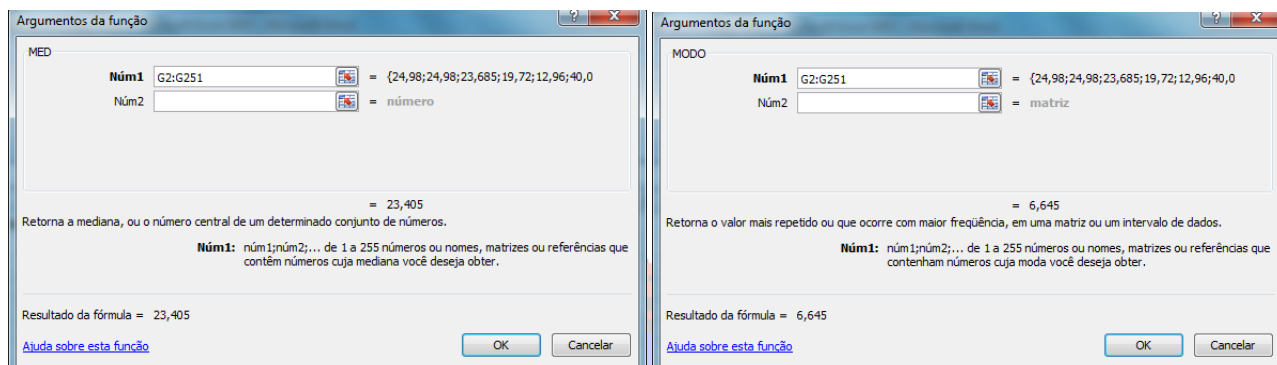
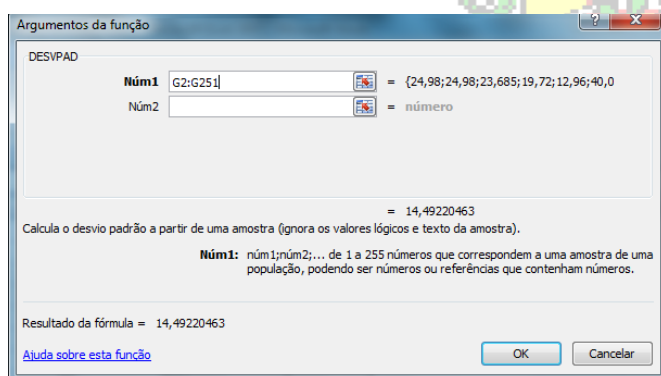


Figura 77 - Argumentos das funções Mediana e Moda (Renda) - 1a versão



É importante ressaltar que o arquivo AmostraToyord.xlsx refere-se a uma AMOSTRA, então devemos calcular o desvio padrão amostral com $n-1$ no denominador da expressão. No Excel isso é feito através da função DESVPAD(). Se houver interesse no desvio padrão populacional, com n no denominador, deve-se usar a função DESVPADP.

Figura 78 - Argumentos da função desvio padrão (Renda) - 1a versão

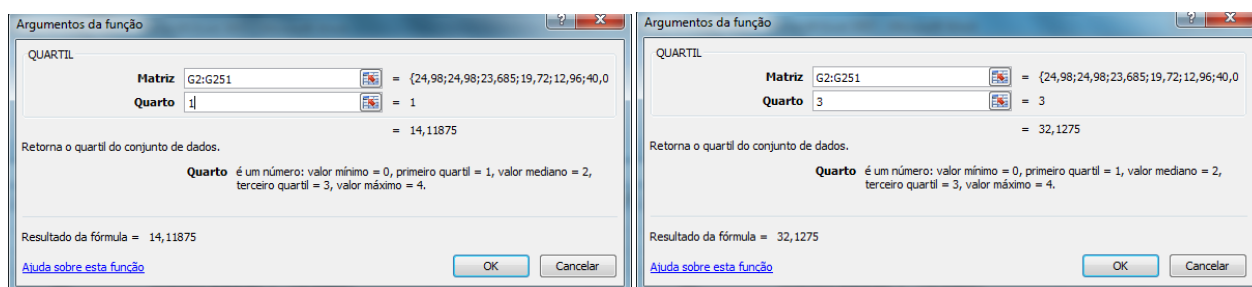


Figura 79 - Argumentos das funções Quartil Inferior e Superior (Renda) - 1a versão

Quando se trata do quartil inferior o Excel exige que se escreva 1 no campo “Quarto”, e 3 para o quartil superior, como explicado na própria função.

² A célula G1 é excluída por conter o nome da variável.

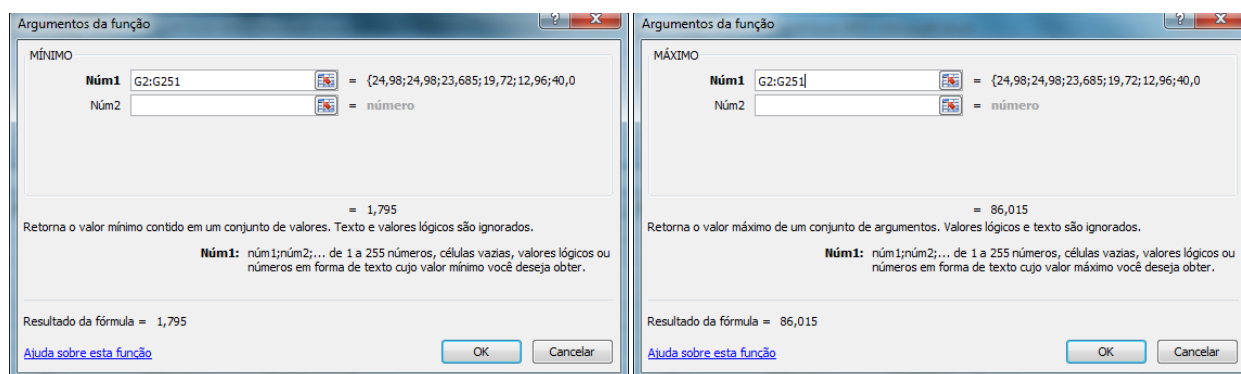


Figura 80 - Argumentos das funções Mínimo e Máximo (Renda) - 1a versão

E os resultados completos podem ser vistos na Figura 81.

Medidas	
Média	25,0622
Mediana	23,405
Moda	6,645
Desvio padrão	14,4922
Quartil inferior	14,1188
Quartil superior	32,1275
Mínimo	1,795
Máximo	86,015

Figura 81 - Medidas de síntese de Renda

A descrição de uma variável quantitativa consiste em analisar basicamente quatro aspectos: tendência central, dispersão, assimetria e presença de valores discrepantes. Para a descrição ser completa precisamos calcular outras medidas de síntese, derivadas das obtidas na Figura 81: coeficiente de variação percentual, diferenças entre os quartis e a mediana, e os valores que definem os valores discrepantes. Da Figura 82 à Figura 84 apresentamos as operações:

	M	N	O	P
1	Medidas			
2	Média	25,0622	CV%	$=(N5/N2)*100$
3	Mediana	23,405		
4	Moda	6,645		
5	Desvio padrão	14,4922		
6	Quartil inferior	14,1188		
7	Quartil superior	32,1275		
8	Mínimo	1,795		
9	Máximo	86,015		

O coeficiente de variação percentual é o resultado da divisão do desvio padrão pela média do conjunto de dados. Permite comparar a dispersão entre grupos com médias diferentes.

Figura 82 - Cálculo do CV%

	M	N	O	P
1	Medidas			
2	Média	25,0622	CV%	57,825
3	Mediana	23,405	Qs-Md	$=N7-N3$
4	Moda	6,645		
5	Desvio padrão	14,4922		
6	Quartil inferior	14,1188		
7	Quartil superior	32,1275		
8	Mínimo	1,795		
9	Máximo	86,015		

	M	N	O	P
1	Medidas			
2	Média	25,0622	CV%	57,825
3	Mediana	23,405	Qs-Md	8,7225
4	Moda	6,645	Md-Qi	$=N3-N6$
5	Desvio padrão	14,4922		
6	Quartil inferior	14,1188		
7	Quartil superior	32,1275		
8	Mínimo	1,795		
9	Máximo	86,015		

Figura 83 - Cálculo das diferenças entre os quartis e a mediana

Se a diferença entre o Quartil superior (Qs) e a Mediana (Md) for bastante próxima da diferença entre a Mediana e o Quartil Inferior (Qi) o conjunto de dados referente à variável pode ser

considerado simétrico em relação à tendência central (além disso, média e mediana devem ser próximas também).

	M	N	O	P	Q
1	Medidas				
2	Média	25,0622	CV%	57,825	
3	Mediana	23,405	Qs-Md	8,7225	
4	Moda	6,645	Md-Qi	9,28625	
5	Desvio padrão	14,4922	Qi - 1,5*(Qs-Qi)	=N6-1,5*(N7-N6)	
6	Quartil inferior	14,1188			
7	Quartil superior	32,1275			
8	Mínimo	1,795			
9	Máximo	86,015			

Figura 84 - Cálculo dos limites para valores discrepantes

Valores discrepantes inferiores são todos aqueles abaixo do limite $Q_i - 1,5 \times (Q_s - Q_i)$. E discrepantes superiores todos acima de $Q_s + 1,5 \times (Q_s - Q_i)$. Os resultados completos estão na Figura 85.

	M	N	O	P
1	Medidas			
2	Média	25,0622	CV%	57,825
3	Mediana	23,405	Qs-Md	8,7225
4	Moda	6,645	Md-Qi	9,28625
5	Desvio padrão	14,4922	Qi - 1,5*(Qs-Qi)	-12,8944
6	Quartil inferior	14,1188	Qs+1,5*(Qs-Qi)	59,1406
7	Quartil superior	32,1275		
8	Mínimo	1,795		
9	Máximo	86,015		

Figura 85 - Medidas de síntese para Renda (completas)

E a descrição da variável Renda será mostrada a seguir.

Tendência Central:

- o valor típico de Renda oscila entre 23,405 (mediana - 50% dos clientes abaixo e 50% acima deste valor) e 25,0622 salários mínimos (média);
- ou seja, entre R\$14557,91 e R\$15588,69 (valores de janeiro de 2012);
- usando dados do Brasil (IBGE – PNAD 2008) que constatou uma renda média do brasileiro de R\$ 1036,00 pode-se dizer que os clientes da operadora de celular tem um bom perfil econômico, pois sua tendência central é pelo menos quatorze vezes maior do que a média do Brasil³.
- a moda vale 6,645 salários mínimos, mas para variáveis quantitativas contínuas esta medida não tem muito significado, pois os valores se repetem pouco⁴.

Dispersão:

- a variação total da renda é de 1,795 (mínimo) a 86,015 (máximo) salários mínimos (R\$1116,49 a R\$53501,33 – valores de janeiro de 2012), a maior renda é quase 48 vezes maior do que a menor, indicando grande dispersão;
- a grande dispersão é confirmada pelo alto valor do coeficiente de variação percentual, indicando que o desvio padrão representa 57,825% da média;
- 25% dos clientes têm renda de até 14,1188 (quartil inferior) salários mínimos (R\$8782), e 25% têm renda acima de 32,1275 (quartil superior) salários mínimos (R\$19983,31) – valores de janeiro de 2012;
- como consequência da afirmação anterior, 50% dos clientes têm renda entre 14,1188 e 32,1275 salários mínimos.

Assimetria

³ É FUNDAMENTAL ter algum padrão para comparação, o que foi possível neste caso.

⁴ Usando a função CONT.SE descobre-se que a moda ocorreu apenas 4 vezes em um total de 250 observações.

- média e mediana são diferentes;
- a conclusão acima é confirmada pela diferença entre quartil superior e mediana (8,7225) que é menor do que a entre a mediana e o quartil inferior (9,28625).

Valores discrepantes

- não há discrepantes inferiores, pois precisaria haver renda inferior -12,8944 salários mínimos, o que é impossível;
- há discrepantes superiores, pois há valores acima⁵ de 59,1406 salários mínimos (R\$36785,45 em janeiro de 2012).

2.3.2 – Medidas de síntese para uma variável quantitativa em função de uma qualitativa – por tabela dinâmica

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: posicionar o cursor em alguma célula do banco de dados, menu Inserir, opção Tabela Dinâmica. O Excel 2007 automaticamente selecionará todas as colunas adjacentes à posição do cursor: no nosso caso, na Plan1, células A1 a K251 do arquivo AmostraToyord.xlsx, como mostrado na Figura 1.

Se pressionarmos OK na Figura 1 o Excel 2007 apresenta a Figura 86, com as várias opções de Layout:

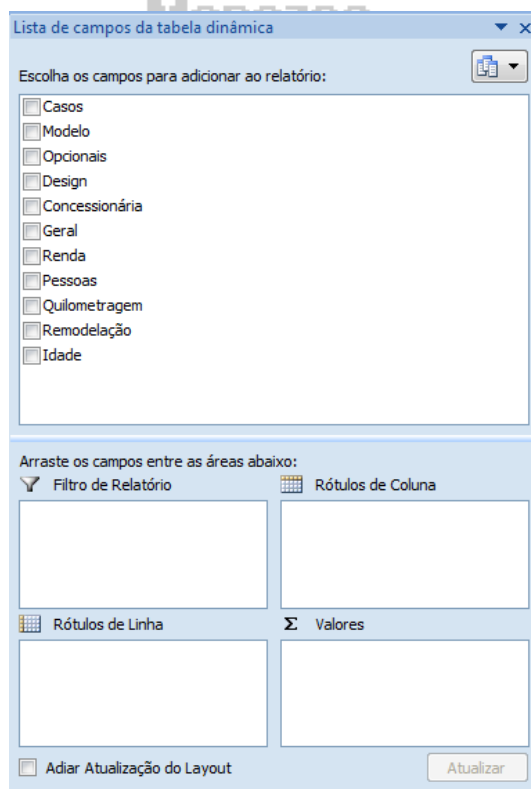


Figura 86 - Layout da tabela dinâmica Renda x Modelo 1a parte

⁵ Usando a função CONT.SE verifica-se que há 7 clientes com renda acima deste valor, sendo portanto valores discrepantes.

Vamos arrastar a variável Modelo para o campo “Rótulos de Linha”, e Renda para o campo “ Σ Valores” (se necessário mude de Soma para Contagem de Renda), NÃO arraste nenhuma outra variável para o campo “Rótulos de Coluna”, resultando na Figura 87.

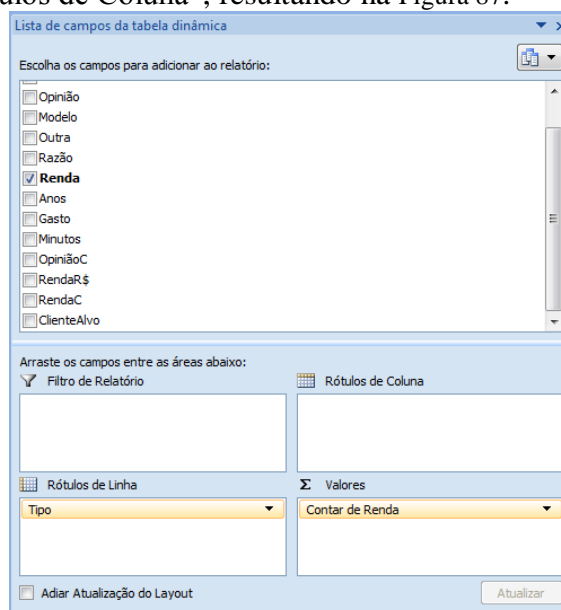


Figura 87 - Layout da tabela dinâmica Renda x Modelo 2a parte

No caso da Figura 87 a tabela apenas contaria quantos clientes há de cada modelo de veículo. Precisamos arrastar a variável Renda mais vezes até o campo “ Σ Valores” e modificar a forma de resumir os resultados: basta selecionar o botão da variável, e escolher “Configurações do campo de Valor”, ver Figura 88.

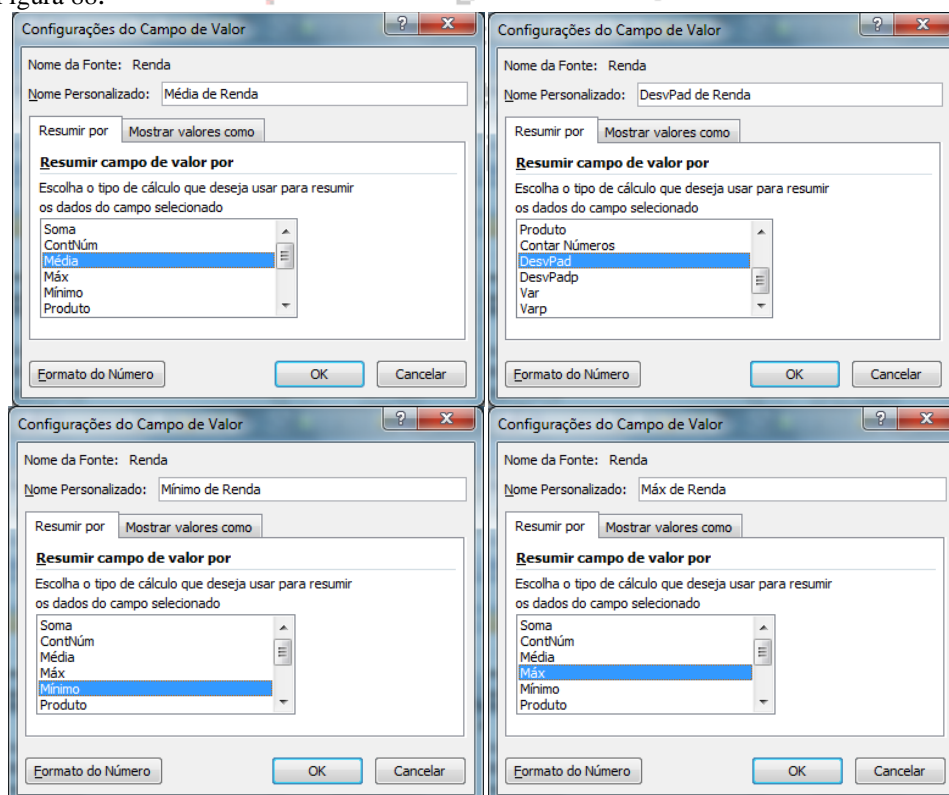


Figura 88 – Formas de resumir os dados de Renda

Usamos “DesvPad” porque os dados do arquivo AmostraToyord.xlsx referem-se a uma amostra. O layout resultante é mostrado na Figura 89– Tela à esquerda. E tal como mostrado na Figura 39, precisamos mover o botão Σ Valores do campo Rótulos de Coluna para o campo Rótulos de Linha, resultando na tela à direita da Figura 89.

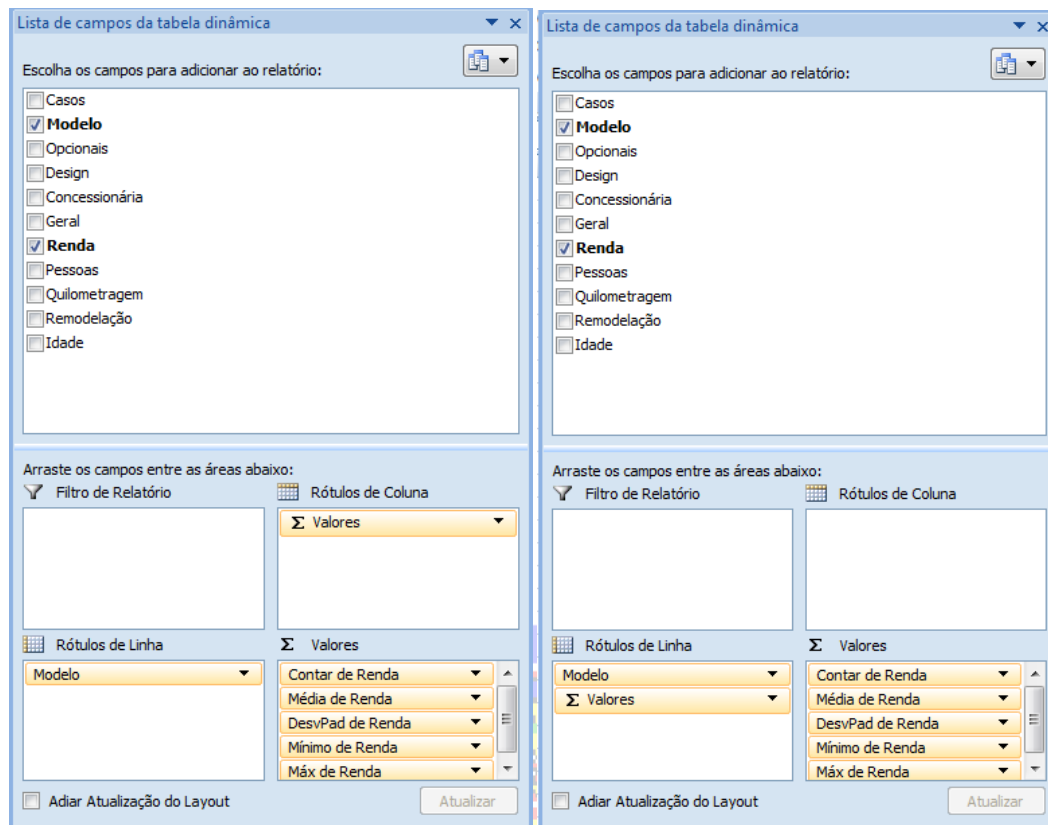


Figura 89 - Layout da tabela dinâmica Renda x Tipo 3a parte

Clicando sobre o X no canto superior direito da tela à direita da Figura 89 chegamos à Figura 90, já sem as células vazias e com modificações para tornar a tabela mais apresentável.

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

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 85, 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.

Figura 90 - Tabela dinâmica de Renda por Tipo

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). Recomenda-se que o cursor seja colocado em uma célula vazia da planilha, para permitir a correta inserção dos dados. No menu “Inserir”, na barra de ferramentas do Excel, em gráficos, escolhe-se dentre estes dispersão e o subtipo inicial, dispersão com marcadores, como mostra a Figura 91.

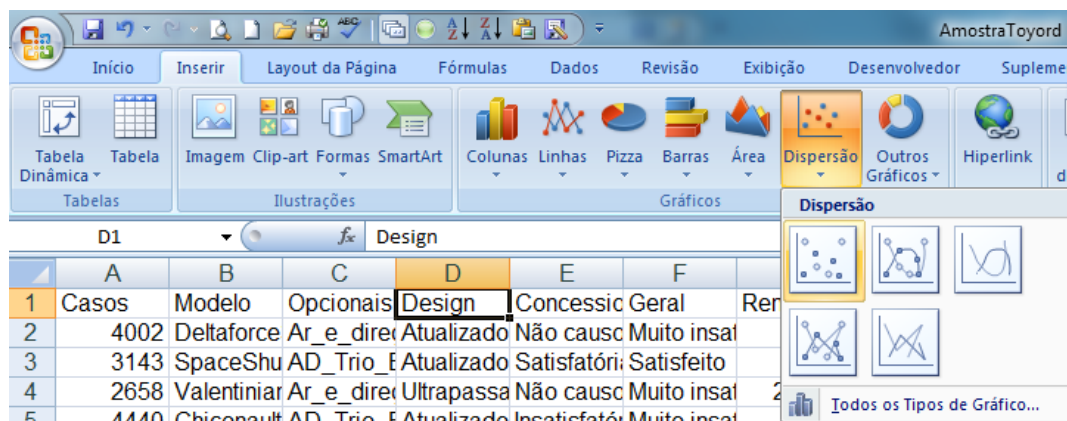


Figura 91 – Inserção de gráfico de dispersão - 1a etapa

Ao selecionar o gráfico na Figura 91 surgirá um gráfico vazio. Selecionando-o e pressionando o botão direito do mouse chegamos à Figura 92.

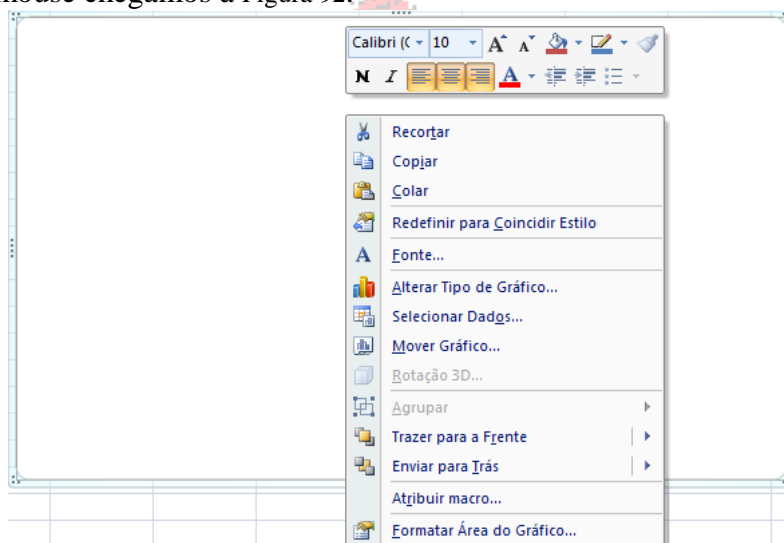


Figura 92- Inserção de gráfico de dispersão - 2a etapa

A opção Selecionar Dados permitirá inserir os dados de Renda e Quilometragem, veja a Figura 93.

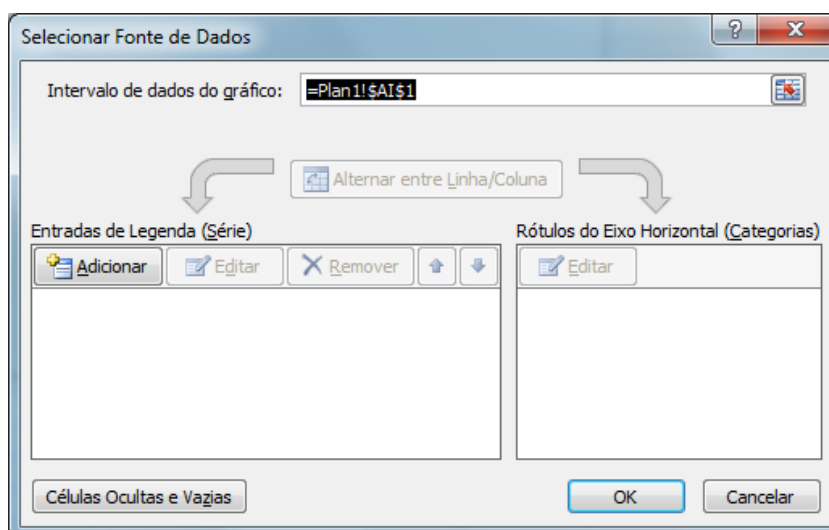


Figura 93 - Adição de sequências - início

Já que não há dados selecionados previamente precisamos pressionar “Adicionar”, o que resultará na Figura 94.

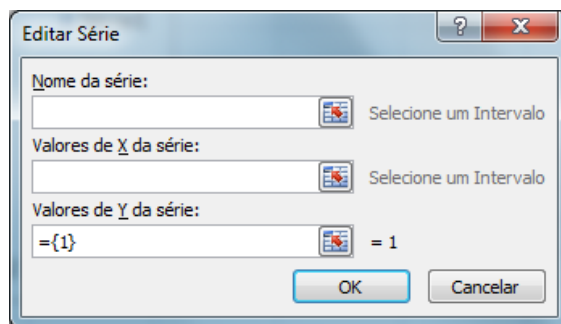


Figura 94 - Adição de sequências – X e Y

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 95.

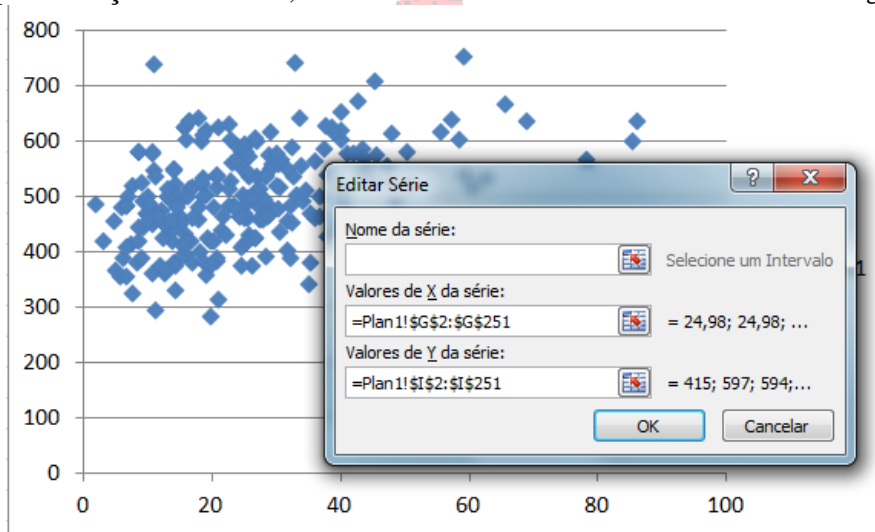


Figura 95 – Adição de sequências - dados inseridos

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. Pressionando OK na Figura 95, chega-se ao gráfico da **Erro! Fonte de referência não encontrada.**

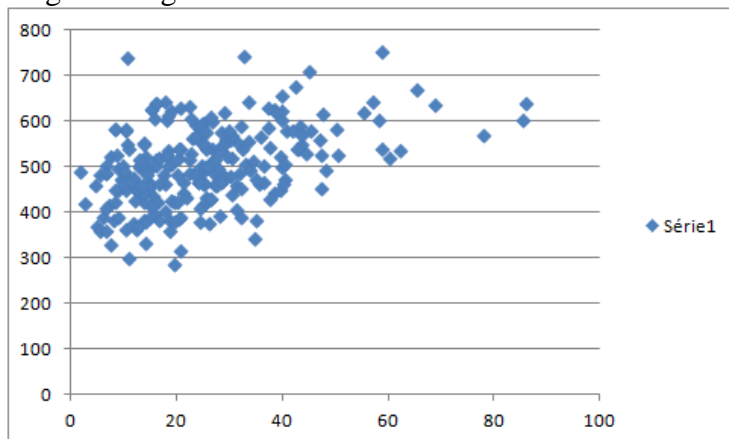


Figura 96 - Diagrama de dispersão de quilometragem por renda – 1 versão

Repare que há uma legenda desnecessária, não há título para o gráfico, não há identificação nos eixos e a escala do eixo Y precisa ser modificada. Estes procedimentos foram vistos na seção 1.1, da Figura 20 em diante, para gráfico por colunas, sendo semelhantes para o gráfico de dispersão. O resultado final então será o apresentado na Figura 97.

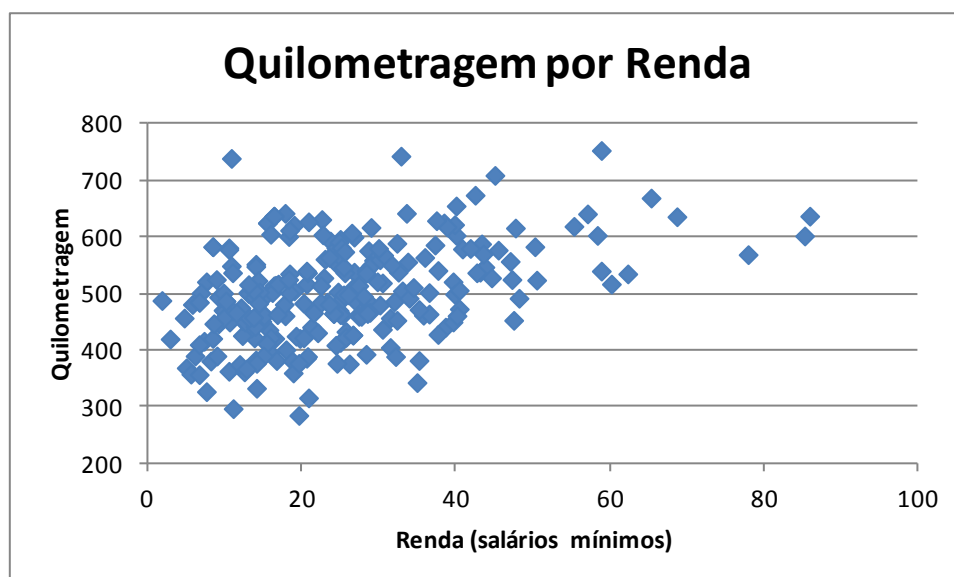


Figura 97 - Diagrama de dispersão de Quilometragem por Renda

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.