

INE 7001 - Procedimentos de Análise Exploratória de Dados utilizando o Excel 2003
 Professor Marcelo Menezes Reis

Índice Analítico

1. Filtragem.....	5
1.1 – Filtragem para identificação de células vazias.....	5
1.2 – Filtragem para correção de erros em variáveis qualitativas	8
1.3 - Filtragem para variáveis quantitativas	9
2. Classificação dos Dados	11
3. Recodificação dos dados.....	14
3.1 – Recodificação de Variáveis Qualitativas	14
3.2 – Recodificação de Variáveis Quantitativas	16
4. Análise de dados de variáveis qualitativas – com tabelas dinâmicas.....	21
4.1 – Uma variável qualitativa	21
4.2 – Duas variáveis qualitativas	27
4.3 – Três Variáveis Qualitativas	36
4.4 – Gráficos para variáveis qualitativas	39
4.4.1 – Gráficos para uma variável qualitativa	39
4.4.2 – Gráficos para duas variáveis qualitativas.....	46
5. Análise de dados de variáveis quantitativas.....	50
5.1 - Procedimentos para variáveis quantitativas discretas – tabelas dinâmicas.....	50
5.1.1 – Distribuição de frequência acumulada.....	55
5.1.2 – Histograma para variáveis quantitativas discretas	56
5.2 - Procedimentos para variáveis quantitativas contínuas.....	61
5.2.1 – Histograma para variáveis quantitativas contínuas.....	67
5.3 – Medidas de síntese	68
5.3.1 – Medidas de síntese para uma variável quantitativa	68
5.3.2 – Medidas de síntese para uma variável quantitativa em função de uma qualitativa – por classificação.....	72
5.3.3 – Medidas de síntese para uma variável quantitativa em função de uma qualitativa – por tabela dinâmica	75

Índice de Figuras

Figura 1 - Arquivo Celula01.xls.....	5
Figura 2 - Seleção de variável para filtragem	6
Figura 3 - Menu Dados: Filtrar	6
Figura 4 - Opções de Filtragem.....	7
Figura 5 - Variável selecionada para filtragem	7
Figura 6 - Valores da variável Tipo - obtidos por Filtragem.....	7
Figura 7 - Variável Tipo - Apenas células Vazias.....	8
Figura 8 - Seleção de um valor da variável Tipo.....	8
Figura 9 - Correção de célula com erro de registro Figura 10 - Células com erros de registro corrigidos	9
Figura 11 - Valores da variável Tipo - sem erros.....	9
Figura 12 - Auto-filtro da variável Anos	10
Figura 13 - Filtro de Renda - Menores Valores Figura 14 - Filtro de Renda - Maiores valores.....	10
Figura 15 - Dados sobre o cliente com a maior renda	10
Figura 16 - Filtragem de Renda - dados para 1,005 salários mínimos.....	11
Figura 17 - Menu Dados - Opção Classificar.....	11
Figura 18 - Opção de classificação apenas por Tipo	12
Figura 19 - Arquivo de dados classificado em função da variável Tipo (parcial)	12
Figura 20 - Opção de classificação por Tipo e Opinião.....	13
Figura 21 - Arquivo de dados classificado em função das variáveis Tipo e Opinião (parcial).....	13
Figura 22 - Criação da variável OpiniãoC	14
Figura 23 - Função SE para recodificação de Opinião	15
Figura 24 - Função SE para recodificar variável Opinião - considerando células vazias.....	16
Figura 25 - Recodificação da variável Opinião - correspondência das células vazias	16
Figura 26 - Função SE para transformar variável Renda - considerando células vazias.....	17
Figura 27 - Transformação da variável Renda - correspondência das células vazias	17
Figura 28 - Estrutura de recodificação da variável Renda	18
Figura 29 - Função de recodificação da variável Renda em RendaC	19
Figura 30 - Filtro da variável recodificada RendaC	19

Figura 31 - Variável ClienteAlvo: resultado da recodificação de OpiniãoC e RendaC	20
Figura 32 - Filtro da variável ClienteAlvo.....	20
Figura 33 - Opções do Menu Dados.....	21
Figura 34 - Etapa 1 da Tabela Dinâmica	21
Figura 35 - Etapa 2 da Tabela Dinâmica	22
Figura 36 - Etapa 2 da Tabela Dinâmica – Dados selecionados.....	22
Figura 37 - Etapa 3 da Tabela Dinâmica – planilha existente.....	22
Figura 38 - Etapa 3 da Tabela Dinâmica - nova planilha.....	23
Figura 39 - Layout da tabela dinâmica	23
Figura 40 - Layout da tabela dinâmica para Tipo.....	23
Figura 41 - Opções de Campo da Tabela Dinâmica	24
Figura 42 - Layout da tabela dinâmica de Tipo: dois campos.....	24
Figura 43 - Mudando a forma de mostrar os dados	25
Figura 44 - Tabela dinâmica de Tipo 1a parte.....	25
Figura 45 - Distribuição de frequências de Tipo.....	26
Figura 46 - Opções de apresentação de Tipo	26
Figura 47 - Tabela dinâmica de Tipo: remoção de células vazias	27
Figura 48 - Tabela dinâmica de Tipo: frequências absolutas e relativas.....	27
Figura 49 – Etapa 2 da Tabela Dinâmica.....	27
Figura 50 - Mensagem de alerta do Excel sobre tabela dinâmica existentes com os mesmos dados	28
Figura 51 - Relatórios de tabela dinâmica disponíveis para os mesmos dados	28
Figura 52 - Etapa 3 da tabela dinâmica.....	28
Figura 53 - Layout da tabela dinâmica Tipo x Opinião 1a parte	29
Figura 54 - Layout da tabela dinâmica Tipo x Opinião 2a parte	29
Figura 55 - Layout da tabela dinâmica Tipo x Opinião 3a parte	30
Figura 56 - Opções de percentuais por linhas	30
Figura 57 - Opções de percentuais por colunas e total	31
Figura 58 - Tabela dinâmica de Tipo x Opinião 1a parte.....	31
Figura 59 - Tabela dinâmica de Tipo x Opinião 2a parte.....	32
Figura 60 - Tabela dinâmica de Tipo x Opinião 3a parte.....	32
Figura 61 - Tabela dinâmica de Tipo x Opinião 4a parte.....	33
Figura 62 - Tabela dinâmica de Tipo x Opinião (sem células vazias).....	33
Figura 63 - Tabela dinâmica Tipo x Opinião - opções de formatação.....	33
Figura 64 - Tabela dinâmica Tipo x Opinião - mudança de ordem de valores.....	34
Figura 65 - Tabela dinâmica de Tipo x Opinião – 1ª versão.....	34
Figura 66 - Tabela dinâmica de Tipo x Opinião – 2ª versão.....	35
Figura 67 - Percentuais de Opinião por Tipo de serviço.....	36
Figura 68 - Tabela dinâmica de Tipo por Opinião em função do Sexo dos clientes.....	37
Figura 69 - Opções de exibição da variável Sexo	37
Figura 70 - Tipo x Opinião para o sexo Feminino.....	37
Figura 71 - Tipo x Opinião para o sexo Masculino.....	37
Figura 72 - Percentuais de Opinião x Sexo	38
Figura 73 - Percentuais de Opinião x Sexo – Pós-pago.....	38
Figura 74 - Percentuais de Opinião x Sexo – Pré-pago	38
Figura 75 - Percentuais de Opinião x Tipo – Feminino.....	38
Figura 76 - Percentuais de Opinião x Tipo –Masculino	39
Figura 77 – Opções de tabela dinâmica: Gráfico dinâmico	39
Figura 78 - Gráfico dinâmico de Tipo - inicial	40
Figura 79 - Gráfico dinâmico de Tipo - Opções de apresentação de Dados	40
Figura 80 - Gráfico dinâmico de Tipo de serviço de telefonia celular - apenas frequências.....	41
Figura 81 - Gráfico dinâmico de Tipo de serviço de telefonia celular - opções	41
Figura 82 - Gráfico dinâmico em colunas de Tipo de serviço de telefonia celular - sem campos de gráfico dinâmico....	42
Figura 83 - Gráfico dinâmico de Tipo de serviço de telefonia celular - opções de gráfico dinâmico.....	42
Figura 84 - Opções de gráfico.....	43
Figura 85 - Gráfico de colunas de Tipo de serviço - Final.....	43
Figura 86 - Gráfico em colunas – opções de eixo	44
Figura 87 - Gráfico em colunas – escala	44
Figura 88 - Mudança de escala no eixo vertical	44
Figura 89 - Gráfico de colunas de Tipo de serviço – escala vertical modificada	44
Figura 90 - Modificação no tipo de gráfico dinâmico	45
Figura 91 - Gráfico dinâmico em pizza de Tipo de serviço de telefonia celular - sem campos de gráfico dinâmico	45
Figura 92 - Opções de Gráfico - gráfico em pizza.....	46
Figura 93 - Gráfico em pizza de Tipo de Serviço - Final.....	46
Figura 94 - Tabela dinâmica de Tipo por Opinião - apenas frequências absolutas	46
Figura 95 - Gráfico de colunas 2D empilhadas de Tipo por Opinião	47
Figura 96 - Mudança de tipo de gráfico de colunas: empilhadas para agrupadas.....	47

Figura 97 - Gráfico de colunas 2D agrupadas de Tipo por Opinião	48
Figura 98 - Gráfico de colunas 2D agrupadas de Tipo por Opinião – Pouco relacionamento.....	48
Figura 99 - Mudança de tipo de gráfico de colunas: agrupadas para 100% empilhadas.....	49
Figura 100 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião.....	49
Figura 101 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião – Pouco relacionamento.....	50
Figura 102 - Layout da tabela dinâmica de Anos.....	51
Figura 103 - Layout da Tabela dinâmica de Anos.....	51
Figura 104 - Mudando a forma de mostrar os dados	51
Figura 105 - Tabela dinâmica de Anos.....	52
Figura 106 – Opções de apresentação de Dados Figura 107 - Tabela de Anos - sem %	52
Figura 108 - Cópia das células da Tabela dinâmica de Anos.....	52
Figura 109 - Inserção de células na Tabela de Anos - 1a parte.....	53
Figura 110 - Menu Inserir - Opção Células	53
Figura 111 - Inserção de células: deslocamento para baixo.....	53
Figura 112 - Células inseridas na Tabela de Anos	54
Figura 113 - Tabela de frequências de Anos.....	54
Figura 114 - Tabela de Anos: frequências e percentuais	54
Figura 115 - CONT.SE: primeira frequência acumulada da variável Ano.....	55
Figura 116 - Tabela de Anos: frequências absolutas, acumuladas e %	56
Figura 117 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa.....	56
Figura 118 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – seleção de valores	57
Figura 119 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – valores adicionados	57
Figura 120 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – seleção de rótulos do eixo X.....	58
Figura 121 - Assistente de gráfico 3a etapa - gráfico de variável quantitativa – Título.....	58
Figura 122 - Histograma de Anos como clientes	58
Figura 123 - Histograma de Anos - opções das colunas.....	59
Figura 124 - Opções de formatar seqüência de dados	59
Figura 125 - Opções de Formatar seqüência de dados.....	60
Figura 126 - Histograma de Anos completos como cliente	60
Figura 127 - Mínimo e máximo para a variável Renda.....	61
Figura 128 - Intervalo para Renda.....	61
Figura 129 - Amplitude de classes de Renda	62
Figura 130 – Cálculo dos limites inferior e do limite superior da primeira classe de Renda	62
Figura 131 – Cálculo dos limites inferior e superior da segunda classe de Renda	63
Figura 132 - Classes de Renda.....	63
Figura 133 - Ponto médio da primeira classe de Renda.....	64
Figura 134 - Obtenção da frequência da primeira classe de Renda.....	64
Figura 135 - Obtenção da frequência da última classe de Renda	65
Figura 136 - Obtenção da frequência acumulada da primeira classe de Renda.....	65
Figura 137 - Obtenção da frequência acumulada da primeira classe de Renda.....	66
Figura 138 - Renda agrupada em classes: frequências absolutas, percentuais, frequências acumuladas e percentuais acumulados	66
Figura 139 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – valores adicionados	67
Figura 140 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – seleção de rótulos do eixo X.....	67
Figura 141 - Histograma da Renda dos clientes em salários mínimos com limites das classes	68
Figura 142 - Fórmulas no Excel 2007	68
Figura 143 - Funções disponíveis no Excel	69
Figura 144 - Funções estatísticas no Excel	69
Figura 145 - Argumentos da função Média (para Renda).....	69
Figura 146 - Argumentos das funções Mediana e Moda (Renda)	70
Figura 147 - Argumentos da função desvio padrão (Renda).....	70
Figura 148 - Argumentos das funções Quartil Inferior e Superior (Renda).....	70
Figura 149 - Argumentos das funções Mínimo e Máximo (Renda)	70
Figura 150 - Medidas de síntese de Renda	71
Figura 151 - Cálculo do CV%.....	71
Figura 152 - Cálculo das diferenças entre os quartis e a mediana	71
Figura 153 - Cálculo dos limites para valores discrepantes	71
Figura 154 - Medidas de síntese para Renda (completas).....	72
Figura 155 - Breakdown de Renda por Tipo - a partir do arquivo classificado.....	73
Figura 156 - Diagrama em caixas de Renda por Tipo	74
Figura 157 - Etapa 3 da tabela dinâmica – Renda por Tipo	75
Figura 158 - Layout da tabela dinâmica Renda x Tipo 1a parte.....	75
Figura 159 – Formas de resumir os dados de Renda	76

Figura 160 - Layout da tabela dinâmica Renda x Tipo 2a parte.....	76
Figura 161 - Tabela dinâmica de Renda por Tipo.....	76
Figura 162 - Layout da tabela dinâmica Renda x Tipo x Sexo	77
Figura 163 - Tabela dinâmica de Renda x Tipo x Sexo (geral).....	77
Figura 164 - Tabelas dinâmicas de Renda x Tipo x Sexo (Feminino e Masculino)	78



**Marcelo
Menezes
Reis**

INE 7001 - 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 em sala, mas utilizando a planilha eletrônica Excel. Os dados estão no arquivo Celular01.xls¹, disponível na páginas das disciplina: contém as informações sobre 5000 clientes da operadora de celular ATCHIM, totalizando 10 variáveis:

- 1) Tipo de serviço: se o cliente é usuário de celular pré-pago ou pós-pago.
- 2) Sexo do cliente: masculino ou feminino.
- 3) Opinião do cliente sobre a operadora ATCHIM: terrível, sofrível, mediana, confiável, excelente.
- 4) Outra operadora do cliente antes de ser cliente da ATCHIM: Cumbuca, Escuridão, Ferrocom, Mortinho ou Nenhuma (o cliente nunca foi cliente de outra operadora além da ATCHIM).
- 5) Razão declarada para ser cliente da ATCHIM (manter-se cliente ou ter mudado de operadora): maior cobertura, qualidade do sinal, tarifa mais baixa.
- 6) Modelo (marca) do aparelho celular: Ericsson, Gradiente, Motorola, Samsung ou Siemens.
- 7) Renda mensal declarada pelo cliente em número de salários mínimos.
- 8) Número de Anos completos em que é cliente da ATCHIM.
- 9) Gasto médio mensal do cliente em reais nos últimos 12 meses, ou no período em que é cliente se inferior a um ano (obtido da base de dados da própria ATCHIM).
- 10) Minutos locais médios do cliente nos últimos 12 meses, ou no período em que é cliente se inferior a um ano (obtido da base de dados da própria ATCHIM).

1. Filtragem

1.1 – Filtragem para identificação de células vazias

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

Acesse o arquivo Celular01.xls, que contém os dados dos clientes de uma fictícia operadora de telefonia celular, a ATCHIM. É possível ver parte do arquivo na figura 1 a seguir:

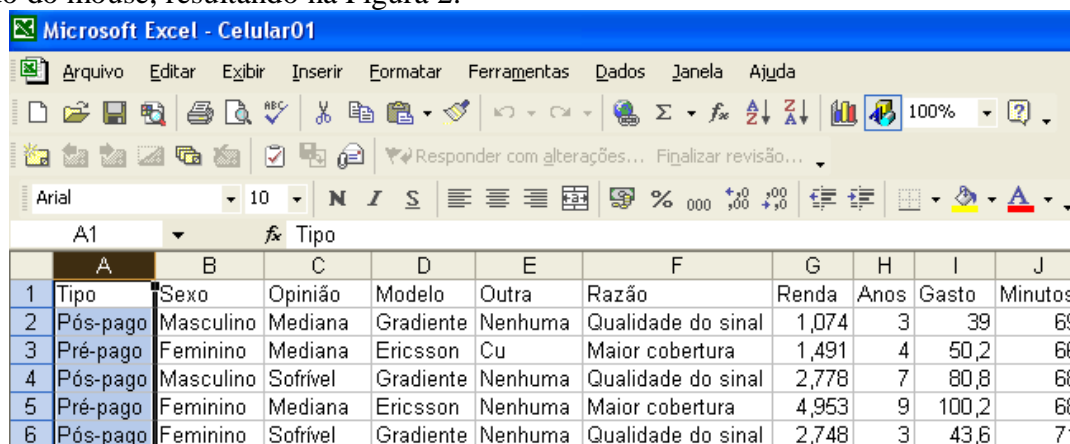
	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69
3	Pré-pago	Feminino	Mediana	Ericsson	Cu	Maior cobertura	1,491	4	50,2	66
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,778	7	80,8	68
5	Pré-pago	Feminino	Mediana	Ericsson	Nenhuma	Maior cobertura	4,953	9	100,2	68
6	Pós-pago	Feminino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,748	3	43,6	71

Figura 1 - Arquivo Celula01.xls

¹ Há também o Celular01.xlsx, para os procedimentos usando o Excel 2007.

Observe as variáveis nas colunas e os dados nas linhas.

Imagine que queremos estudar a variável **qualitativa** Tipo, usando filtragem. O primeiro passo consiste em selecionar a variável, pondo o cursor sobre a coluna e pressionando o botão esquerdo do mouse, resultando na Figura 2.

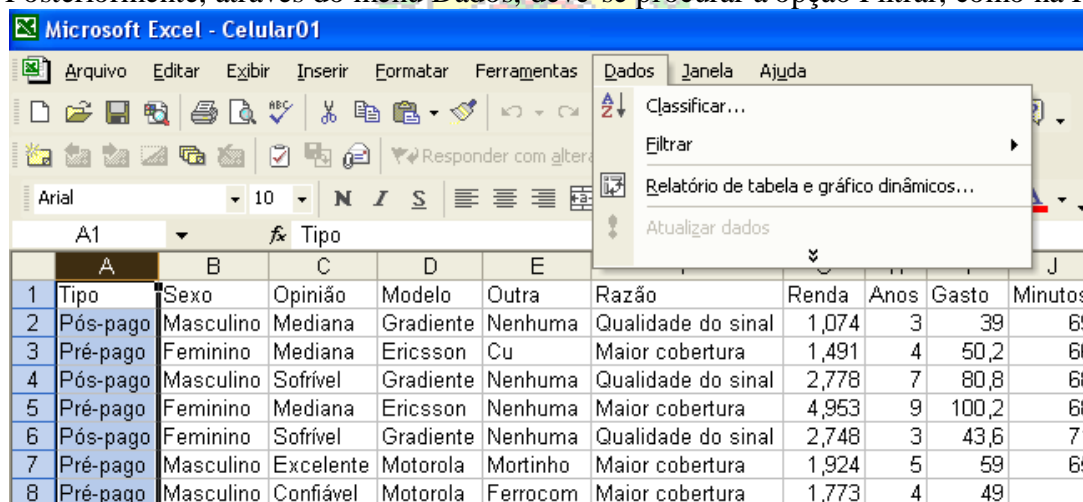


The screenshot shows the Microsoft Excel interface with the 'Dados' menu open. The 'Tipo' column (A1) is selected. The data table is as follows:

	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69
3	Pré-pago	Feminino	Mediana	Ericsson	Cu	Maior cobertura	1,491	4	50,2	66
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,778	7	80,8	68
5	Pré-pago	Feminino	Mediana	Ericsson	Nenhuma	Maior cobertura	4,953	9	100,2	68
6	Pós-pago	Feminino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,748	3	43,6	71

Figura 2 - Seleção de variável para filtragem

Posteriormente, através do menu Dados, deve-se procurar a opção Filtrar, como na Figura 3.



The screenshot shows the 'Dados' menu open with the 'Filtrar' option selected. The data table is as follows:

	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69
3	Pré-pago	Feminino	Mediana	Ericsson	Cu	Maior cobertura	1,491	4	50,2	66
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,778	7	80,8	68
5	Pré-pago	Feminino	Mediana	Ericsson	Nenhuma	Maior cobertura	4,953	9	100,2	68
6	Pós-pago	Feminino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,748	3	43,6	71
7	Pré-pago	Masculino	Excelente	Motorola	Mortinho	Maior cobertura	1,924	5	59	65
8	Pré-pago	Masculino	Confiável	Motorola	Ferocom	Maior cobertura	1,773	4	49	

Figura 3 - Menu Dados: Filtrar

Dentro da opção Filtrar há duas escolhas possíveis: o Auto-Filtro e o Filtro Avançado (ver Figura 4). No Auto-Filtro o Excel automaticamente seleciona todos os valores possíveis para a variável, no Filtro Avançado é possível definir critérios (em células do Excel) para orientar a filtragem.

Vamos usar a opção Auto-Filtro, que uma vez selecionada acrescenta uma seta à coluna da variável sob filtragem, para que seja possível identificá-la, veja a Figura 4.

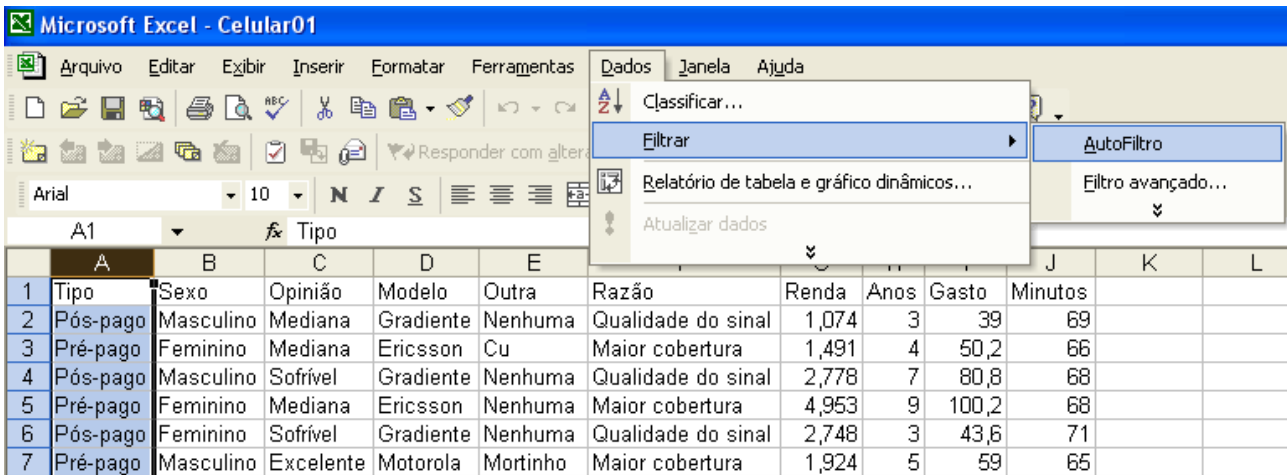
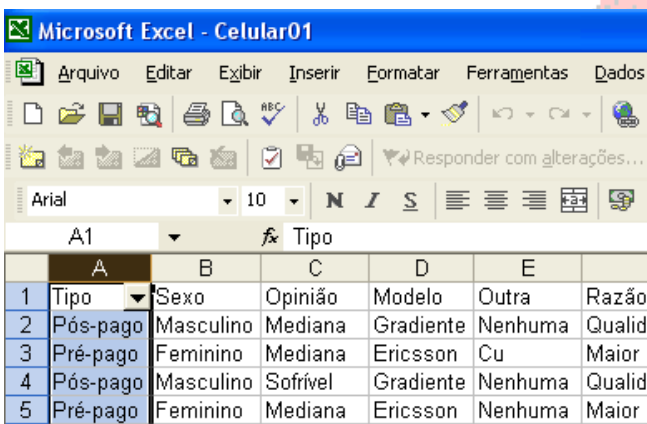


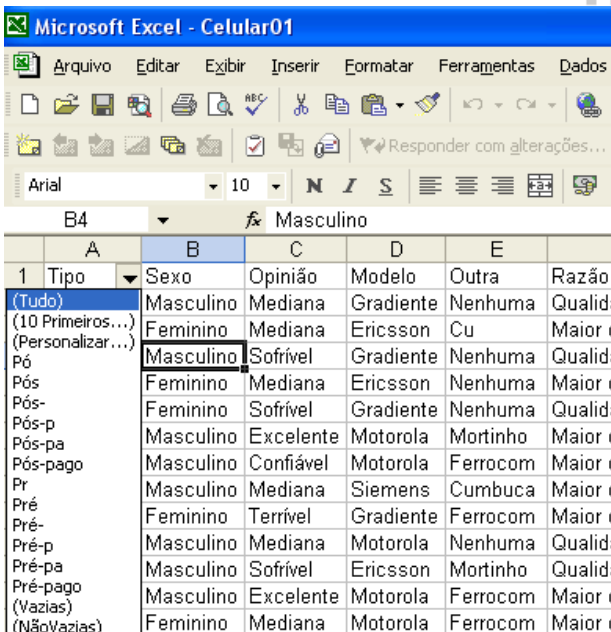
Figura 4 - Opções de Filtragem



Se pressionarmos a seta ao lado da variável Tipo (na Figura 5) o Excel apresenta uma lista com todos os valores registrados da variável, incluindo células vazias e eventuais erros.

O resultado pode ser observado na Figura 6.

Figura 5 - Variável selecionada para filtragem



Os valores corretos para a variável Tipo são Pré-pago e Pós-pago. Mas, há vários registros incorretos para ambos, em que houve truncamento de letras. É possível ver também as opções: “Vazias”, se selecionada permitirá avaliar a quantidade de dados perdidos; “10 primeiros”; “Personalizar”, na qual podemos construir o filtro de nosso interesse; “Não Vazias”, se selecionada mostrará apenas as células que contém registros.

Selecionando qualquer dos valores o Excel apresentará apenas as células que o contém. Selecionando “Vazias” obtemos a Figura 7.

Figura 6 - Valores da variável Tipo - obtidos por Filtragem

	A	B	C	D	E
1	Tipo	Sexo	Opinião	Modelo	Outra
52		Feminino	Terrível	Gradiente	Escureidão
605		Feminino	Excelente	Motorola	Cumbuca
1250		Feminino	Confiável	Siemens	Nenhuma
2414		Feminino	Mediana	Motorola	Escureidão
2638		Feminino	Mediana	Samsung	Escureidão
4051		Masculino	Sofrível	Samsung	Ferocom
4192		Masculino	Sofrível	Gradiente	Nenhuma

Figura 7 - Variável Tipo - Apenas células Vazias

Este procedimento precisa ser feito para todas as variáveis (qualitativas e quantitativas).

Outro uso da Filtragem é a identificação de erros de registro (ver Figura 6), e a sua correção.

1.2 – Filtragem para correção de erros em variáveis qualitativas

Basta selecionar um dos valores incorretamente registrados, como “Pó” na Figura 8.

	A	B	C	D	E
1	Tipo	Sexo	Opinião	Modelo	Outra
	(Tudo)	Feminino	Terrível	Ericsson	Nenhuma
	(10 Primeiros...)	Feminino	Terrível	Motorola	Nenhuma
	(Personalizar...)	Masculino	Terrível	Siemens	Nenhuma
	Pó	Feminino	Sofrível	Motorola	Mortinho
	Pós				
	Pós-				
	Pós-p				
	Pós-pa				
	Pós-pago				
	Pr				
	Pré				
	Pré-				
	Pré-p				
	Pré-pa				
	Pré-pago				
	(Vazias)				
	(NãoVazias)				

Figura 8 - Seleção de um valor da variável Tipo

O Excel apresentará apenas as células que contém o valor incorreto “Pó”. Na Figura 9 podemos ver que são 4 células apenas. Não obstante, precisamos corrigi-las.

	A	B	C	D	E
1	Tipo	Sexo	Opinião	Modelo	Outra
234	Pós-pago	Feminino	Terrível	Ericsson	Nenhuma
426	Pó	Feminino	Terrível	Motorola	Nenhuma
439	Pó	Masculino	Terrível	Siemens	Nenhuma
3903	Pó	Feminino	Sofrível	Motorola	Mortinho

Figura 9 - Correção de célula com erro de registro

	A	B	C	D	E
1	Tipo	Sexo	Opinião	Modelo	Outra
234	Pós-pago	Feminino	Terrível	Ericsson	Nenhuma
426	Pós-pago	Feminino	Terrível	Motorola	Nenhuma
439	Pós-pago	Masculino	Terrível	Siemens	Nenhuma
3903	Pós-pago	Feminino	Sofrível	Motorola	Mortinho

Figura 10 - Células com erros de registro corrigidos

Basta digitar o valor correto na primeira célula do intervalo (A234 na Figura 9). Em seguida, “arrasta-se” o valor correto até a última célula com o valor incorreto “Pó”, e todos serão corrigidos, como mostrado na Figura 10.

Embora pareça maçante, o procedimento descrito acima precisa ser realizado para todos os valores incorretos, e para todas as variáveis. Para o caso da variável Tipo, uma vez corrigidos todos os erros de registro, ao pressionar a seta do Filtro surgirá a tela da Figura 11:

	A	B	C	D	E
1	Tipo	Sexo	Opinião	Modelo	Outra
	(Tudo)	Feminino	Mediana	Samsung	Nenhuma
	(10 Primeiros...)	Masculino	Mediana	Gradiente	Nenhuma
	(Personalizar...)				
	Pós-pago				
	Pré-pago				
	(Vazias)				
	(Não Vazias)				

Figura 11 - Valores da variável Tipo - sem erros

Restam apenas os valores (corretos) Pós-pago e Pré-pago, e as opções de filtragem: Tudo, 10 primeiros, Personalizar, Vazias e Não Vazias.

Observe que com um procedimento simples é possível corrigir todos os erros de registro, avaliar a quantidade de dados perdidos e eventualmente (através da opção Personalizar) realizar outras filtragens.

1.3 - Filtragem para variáveis quantitativas

Para variáveis quantitativas a filtragem é extremamente útil para identificar os valores extremos do conjunto (eventuais valores discrepantes), além dos dados perdidos.

Se selecionarmos a variável Anos, **quantitativa discreta**, e acionarmos a opção AutoFiltro (menu Dados, opção Filtrar), tal como feito com a variável Tipo nas Figuras 2, 3 e 4, chegamos à tela mostrada na Figura 12.

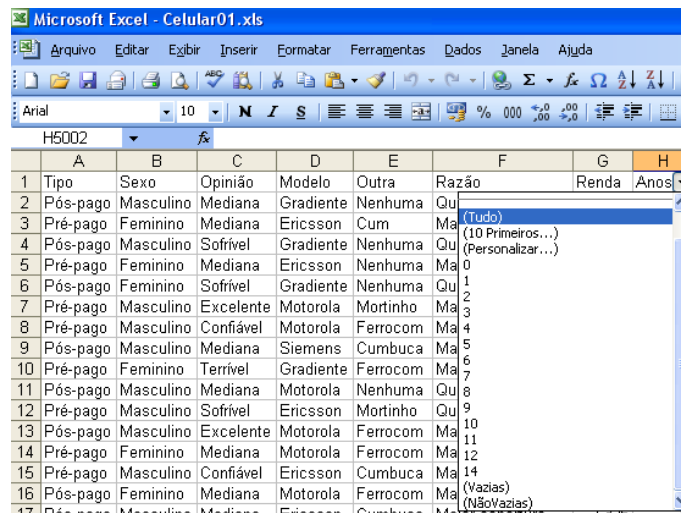


Figura 12 - Auto-filtro da variável Anos

Os valores da variável Anos variam de 0 a 14 anos (há células vazias também). O valor 0 corresponde a pessoas que ainda não completaram um ano inteiro como clientes, e o valor 14 às pessoas que são clientes desde o início dos negócios da operadora. Apenas os valores que ocorrem nos dados são apresentados: o valor 13 é um dos valores possíveis para a variável Anos, mas como não há nenhum cliente registrado na amostra com este valor, ele não é apresentado.

Se realizarmos o procedimento de filtragem com a variável Renda, quantitativa contínua, vamos observar a grande quantidade de valores possíveis entre o mínimo (1,003 salários mínimos), mostrado na Figura 13, e o máximo (43,878 salários mínimos), mostrado na Figura 14.

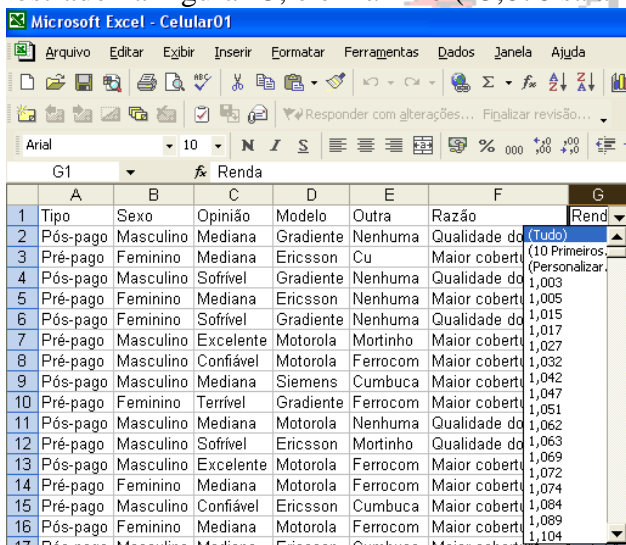


Figura 13 - Filtro de Renda - Menores Valores

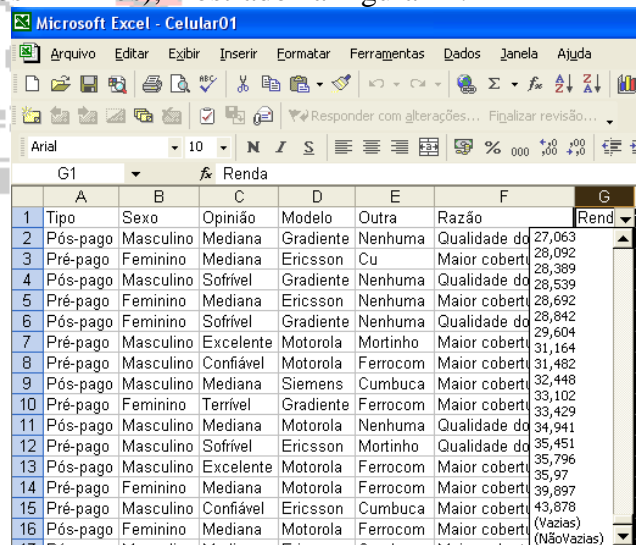


Figura 14 - Filtro de Renda - Maiores valores

Se quisermos saber mais sobre o cliente com a maior renda basta selecionar o valor 43,878 e teremos a tela da Figura 15, com todas as informações disponíveis.

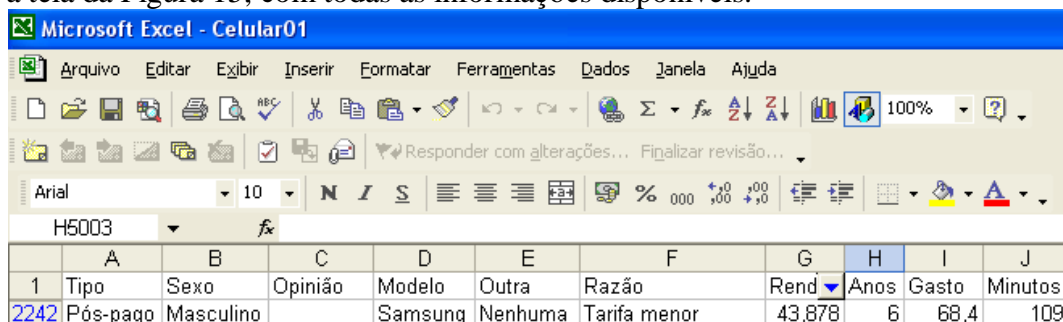
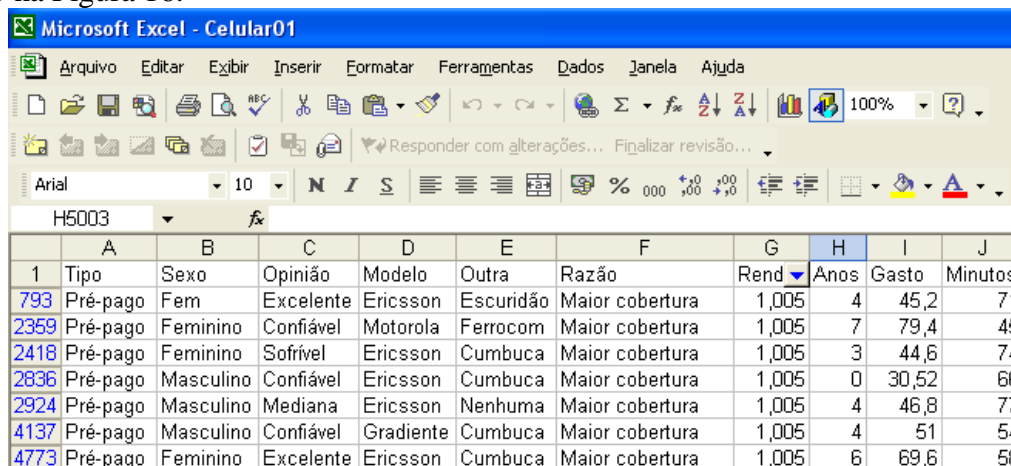


Figura 15 - Dados sobre o cliente com a maior renda

Trata-se de um cliente de serviço pós-pago, do sexo masculino, que optou por um modelo Samsung, não foi cliente de nenhuma operadora anteriormente. Ele declarou que a razão para estar na ATCHIM é a tarifa menor, é cliente há 6 anos, gasta em média R\$ 68,4 mensais, falando 109 minutos. Mas, não há resultado para a variável Opinião.

Poderíamos repetir o procedimento para qualquer outro valor de Renda e conhecer melhor o perfil dos respectivos clientes. Por exemplo, poderíamos escolher o valor 1,005 salários mínimos, resultando na Figura 16.



	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Rend	Anos	Gasto	Minutos
793	Pré-pago	Fem	Excelente	Ericsson	Escuridão	Maior cobertura	1,005	4	45,2	71
2359	Pré-pago	Feminino	Confiável	Motorola	Ferocom	Maior cobertura	1,005	7	79,4	45
2418	Pré-pago	Feminino	Sofrível	Ericsson	Cumbuca	Maior cobertura	1,005	3	44,6	74
2836	Pré-pago	Masculino	Confiável	Ericsson	Cumbuca	Maior cobertura	1,005	0	30,52	66
2924	Pré-pago	Masculino	Mediana	Ericsson	Nenhuma	Maior cobertura	1,005	4	46,8	77
4137	Pré-pago	Masculino	Confiável	Gradiente	Cumbuca	Maior cobertura	1,005	4	51	54
4773	Pré-pago	Feminino	Excelente	Ericsson	Cumbuca	Maior cobertura	1,005	6	69,6	58

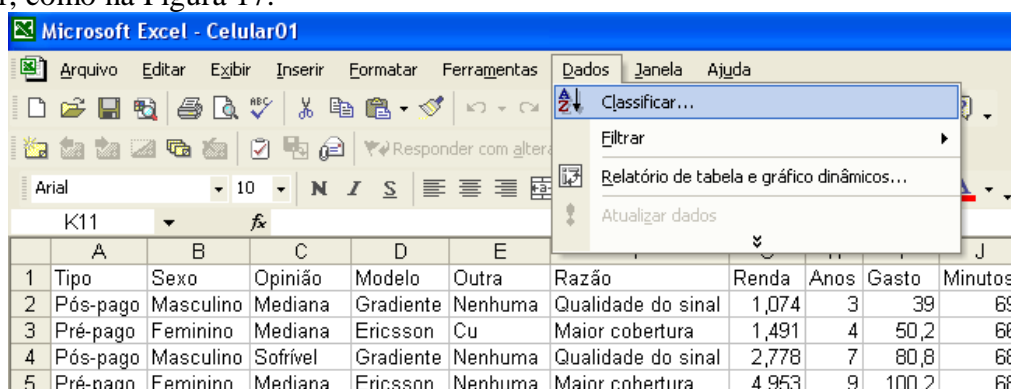
Figura 16 - Filtragem de Renda - dados para 1,005 salários mínimos

Os clientes que percebem 1,005 salários mínimos preferem serviço pré-pago, e estão na ATCHIM pela maior cobertura. Todos são clientes com pelo menos 45 anos, e as demais variáveis têm diferentes valores.

Marcelo
Menezes
Reis

2. Classificação dos Dados

Em muitos casos há interesse em reordenar o conjunto de dados de maneira a facilitar a visualização dos valores específicos de uma ou mais variáveis. Isso será extremamente útil nas próximas aulas quando construirmos tabelas para os dados. O Microsoft Excel dispõe de vários mecanismos automáticos de classificação, que podem ser acessados pelo menu Dados, opção Classificar, como na Figura 17.



	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69
3	Pré-pago	Feminino	Mediana	Ericsson	Cu	Maior cobertura	1,491	4	50,2	66
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,778	7	80,8	68
5	Pré-pago	Feminino	Mediana	Ericsson	Nenhuma	Maior cobertura	4,953	9	100,2	68

Figura 17 - Menu Dados - Opção Classificar

Podemos classificar os dados em função de até 3 variáveis, de forma automática. Se desejássemos fazer a classificação apenas em função de Tipo bastaria escolhê-la na primeira opção da tela da Figura 18.

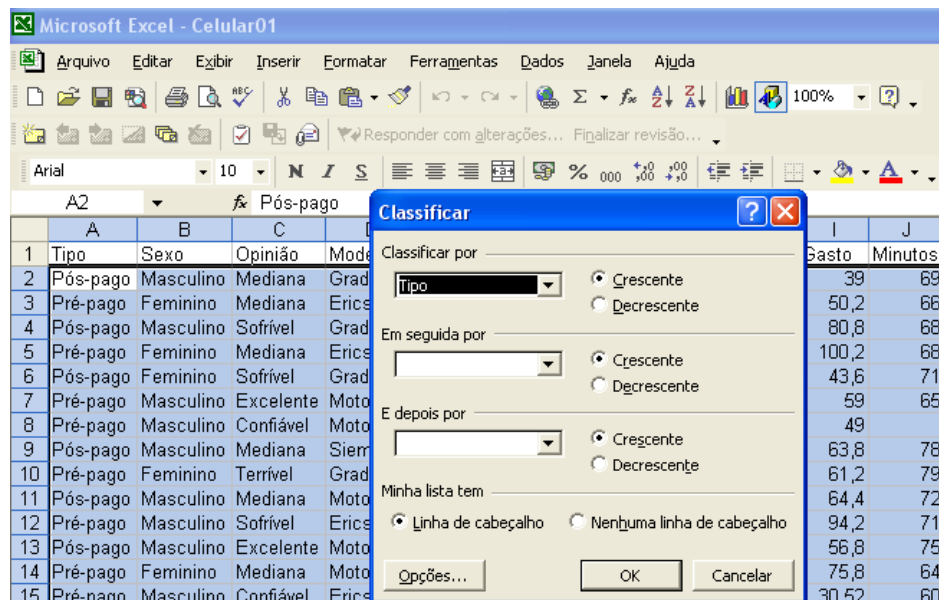


Figura 18 - Opção de classificação apenas por Tipo

Uma vez pressionado “OK”, o resultado será a tela da Figura 19.

	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2		Feminino	Terrível	Gradiente	Escuridão	Maior cobertura	4,725	6	67,8	65
3		Feminino	Excelente	Motorola	Cumbuca	Maior cobertura	1,953	2	33,2	60
4		Feminino	Confiável	Siemens	Nenhuma	Maior cobertura	1,978	2	30,6	56
5		Feminino	Mediana	Motorola	Escuridão	Maior cobertura	1,393	7	82	58
6		Feminino	Mediana	Samsung	Escuridão	Maior cobertura	9,768	2	30,52	70
7		Masculino	Sofrível	Samsung	Ferocom	Maior cobertura	2,145	3	35	55
8		Masculino	Sofrível	Gradiente	Nenhuma	Tarifa menor	15,9	9	104	91
9	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69
10	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,778	7	80,8	68
11	Pós-pago	Feminino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,748	3	43,6	71
12	Pós-pago	Masculino	Mediana	Siemens	Cumbuca	Maior cobertura	6,552	5	63,8	78
13	Pós-pago	Masculino	Mediana	Motorola	Nenhuma	Qualidade do sinal	4,419	5	64,4	72
14	Pós-pago	Masculino	Excelente	Motorola	Ferocom	Maior cobertura	8,967	5	56,8	75

Figura 19 - Arquivo de dados classificado em função da variável Tipo (parcial)

Todo o arquivo foi reordenado em função dos valores da variável Tipo, começando pelas células vazias, depois por Pós-pago e finalmente por Pré-pago (o Excel classifica as variáveis qualitativas por ordem alfabética). O uso da classificação mantém a consistência do banco de dados, pois todos os dados (todas as linhas) são reorganizados mantendo a coerência. Por exemplo, a primeira informação de Pós-pago estava (ver Figura 18) na linha 2, com os valores Masculino para Sexo, Mediana para Opinião, Gradiente para Modelo, etc. Após a classificação as informações são deslocadas para a linha 9 (ver Figura 19), primeira linha do valor Pós-pago de Tipo, logo após a última linha com célula vazia para a variável, mas mantendo a consistência das informações.

Se houvesse interesse em fazer a classificação em função de Tipo, e depois por Opinião, deveríamos repetir os procedimentos das Figuras 17 e 19, mas agora escolhendo Opinião na segunda opção de classificação, tal como mostrado na Figura 20.

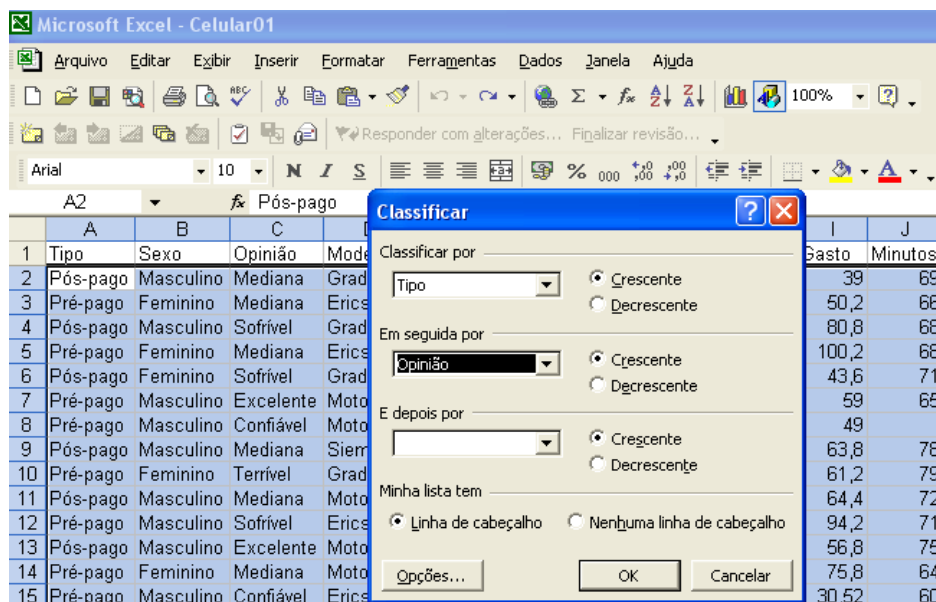


Figura 20 - Opção de classificação por Tipo e Opinião

O arquivo classificado é mostrado na Figura 21.

	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2		Feminino	Confiável	Siemens	Nenhuma	Maior cobertura	1,978	2	30,6	56
3		Feminino	Excelente	Motorola	Cumbuca	Maior cobertura	1,953	2	33,2	60
4		Feminino	Mediana	Motorola	Escuridão	Maior cobertura	1,393	7	82	58
5		Feminino	Mediana	Samsung	Escuridão	Maior cobertura	9,768	2	30,52	70
6		Masculino	Sofrível	Samsung	Ferrocom	Maior cobertura	2,145	3	35	55
7		Masculino	Sofrível	Gradiente	Nenhuma	Tarifa menor	15,9	9	104	91
8		Feminino	Terrível	Gradiente	Escuridão	Maior cobertura	4,725	6	67,8	65
9	Pós-pago	Masculino		Motorola	Nenhuma	Qualidade do sinal	10,773	3	35,4	97
10	Pós-pago	Masculino		Samsung	Nenhuma	Tarifa menor	10,614	4	47	93
11	Pós-pago	Masculino		Samsung	Nenhuma	Tarifa menor	43,878	6	68,4	109
12	Pós-pago	Masculino		Gradiente	Nenhuma	Qualidade do sinal	3,657	5	64	75
13	Pós-pago	Masculino	Confiável	Motorola	Cumbuca	Qualidade do sinal	15,087	1	30,52	91
14	Pós-pago	Masculino	Confiável	Ericsson	Ferrocom	Maior cobertura	2,688	8	91,2	67
15	Pós-pago	Feminino	Confiável	Ericsson	Ferrocom	Maior cobertura	4,725	4	51,4	72
16	Pós-pago	Masculino	Confiável	Motorola	Nenhuma	Qualidade do sinal	4,725	6	148	76
17	Pós-pago	Feminino	Confiável	Gradiente	Nenhuma	Maior cobertura	3,657	4	49,6	57
18	Pós-pago	Feminino	Confiável	Samsung	Escuridão	Maior cobertura	5,688	6	70,4	66

Figura 21 - Arquivo de dados classificado em função das variáveis Tipo e Opinião (parcial)

Observe que para as células vazias (primeiros valores de Tipo), os valores de Opinião são ordenados (também alfabeticamente): há um valor Confiável, um Excelente, dois valores Mediana, dois Sofrível e um Terrível. Em seguida, o próximo valor de Tipo, é Pós-pago, e a variável Opinião é ordenada de acordo, começando pelas células vazias e depois pelas células com Confiável. Todas as outras variáveis são reordenadas em função destas duas.

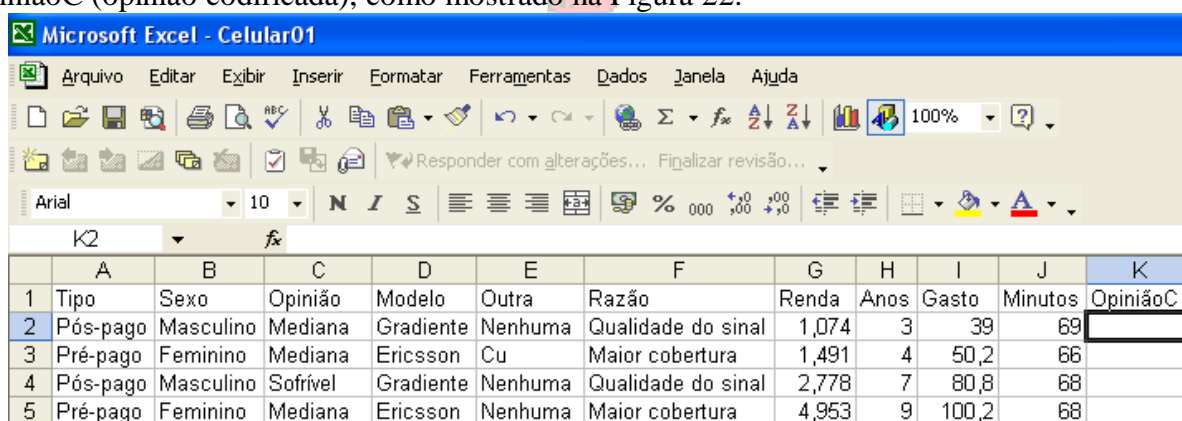
3. Recodificação dos dados

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

3.1 – Recodificação de Variáveis Qualitativas

Vamos recodificar a variável Opinião, agrupando as opiniões positivas (confiável e excelente) e as negativas (Mediana, Sofrível e Terrível), para nas próximas aulas conseguirmos construir tabelas desta variável.

O primeiro passo é seleccionar a célula K1, onde será colocado o nome da nova variável, OpiniãoC (opinião codificada), como mostrado na Figura 22.



	A	B	C	D	E	F	G	H	I	J	K
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69	
3	Pré-pago	Feminino	Mediana	Ericsson	Cu	Maior cobertura	1,491	4	50,2	66	
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,778	7	80,8	68	
5	Pré-pago	Feminino	Mediana	Ericsson	Nenhuma	Maior cobertura	4,953	9	100,2	68	

Figura 22 - Criação da variável OpiniãoC

Para fazer a recodificação da variável nós devemos construir uma fórmula que relaciona os valores da variável Opinião com os novos valores previstos para OpiniãoC. Para construir uma fórmula no Excel posicionamos o cursor na célula desejada, digitamos =, seguida da fórmula. Para o caso da recodificação nós queremos implementar a seguinte fórmula:

SE Opinião = Confiável OU Excelente ENTÃO OpiniãoC = Positiva
SENÃO OpiniãoC = Negativa.

Geralmente a recodificação envolve alguma expressão lógica, como a descrita acima: SE os valores de Opinião forem Confiável, ou Excelente, ENTÃO OpiniãoC será Positiva; SENÃO (se os valores de Opinião forem Mediana, Sofrível ou Terrível), então OpiniãoC será Negativa. A colocação da fórmula correspondente na célula K2 possibilitaria recodificar a variável.

A Figura 23 a seguir ilustra o procedimento que queremos fazer.

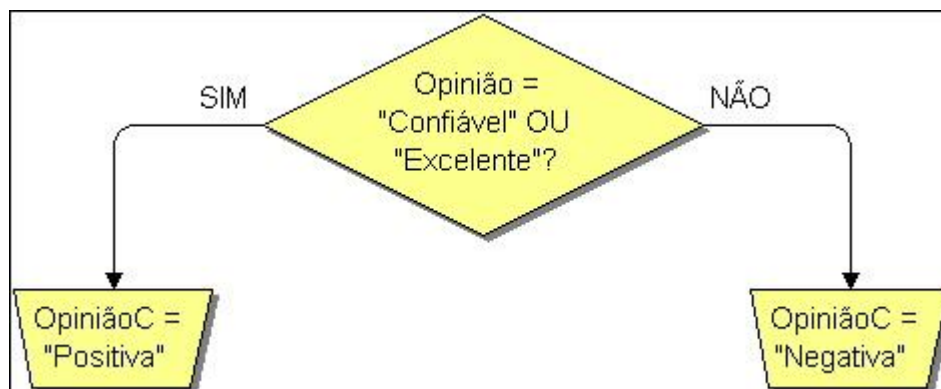


Figura 23 - Função SE para recodificação de Opinião

Observe que os valores de opinião estão entre aspas. No Excel, toda vez que realizamos operações com variáveis qualitativas precisamos pôr seus valores entre aspas, ou causará erro.

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

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

- E(teste lógico1; teste lógico2; ...). Retornará VERDADEIRO se todos os testes lógicos resultarem verdadeiros; retornará FALSO se um ou mais testes lógicos resultarem falsos. Teste lógico1; teste lógico2;... são de 1 a 30 condições para testar e que podem ser VERDADEIRO ou FALSO.

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

Menezes

No caso da recodificação da variável Opinião vamos usar uma função SE e dentro dela, no teste lógico uma função OU: SE Opinião for Confiável OU Excelente, OpiniãoC será Positiva, SENÃO será Negativa. Mas, como há células vazias na variável Opinião, se realizarmos a recodificação tal como mostrada aqui “criaremos” informação: como as células vazias não contém as opiniões Confiável ou Excelente serão registradas em OpiniãoC como “Negativas”, o que não é admissível. Precisamos refinar a função SE para que ela leve em conta a existência das células vazias, o que será possível introduzindo mais uma função SE, da seguinte maneira.

SE Opinião = “ ” ENTÃO OpiniãoC = “ ”

SENÃO SE Opinião = Confiável OU Excelente ENTÃO OpiniãoC = Positiva

SENÃO OpiniãoC = Negativa.

Ou, em termos da sintaxe do Excel:

= SE(C2=“ ”;C2;SE(OU(C2=“Excelente”;C2=“Confiável”);“Positiva”;“Negativa”)).

Observe cuidadosamente o posicionamento dos parênteses: o teste que vimos originalmente ficou “embutido” na ação a ser feita se o primeiro teste (se a célula de Opinião é vazia) resultar falso. Veja a implementação na Figura 24.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC				
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69	=SE(C2="" ; C2; SE(OU(C2="Excelente"; C2="Confiável"); "Positiva"; "Negativa"))				
3	Pré-pago	Feminino	Mediana	Ericsson	Cu	Maior cobertura	1,491	4	50,2	66	"Positiva"; "Negativa"))				
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	2,778	7	80,8	68	Negativa				
5	Pré-pago	Feminino	Mediana	Ericsson	Nenhuma	Maior cobertura	4,953	9	100,2	68	Negativa				

Figura 24 - Função SE para recodificar variável Opinião - considerando células vazias

Arrastando a fórmula até a célula K5001 a variável estará corretamente recodificada. Se repetirmos a filtragem na variável Opinião, escolhendo a exibição das células vazias, o resultado será a Figura 25.

	A	B	C	D	E	F	G	H	I	J	K
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC
1103	Pós-pago	Masculino		Motorola	Nenhuma	Qualidade do sinal	10,773	3	35,4	97	
1631	Pós-pago	Masculino		Samsung	Nenhuma	Tarifa menor	10,614	4	47	93	
2156	Pré-pago	Feminino		Gradiente	Cumbuca	Maior cobertura	1,888	0	30,52	48	
2187	Pré-pago	Feminino		Ericsson	Cumbuca	Maior cobertura	1,074	2	30,52	42	
2242	Pós-pago	Masculino		Samsung	Nenhuma	Tarifa menor	43,878	6	68,4	109	
2323	Pré-pago	Masculino		Gradiente	Cumbuca	Maior cobertura	1,42	3	38,8	67	
3095	Pré-pago	Feminino		Gradiente	Escuridão	Maior cobertura	2,871	7	78,4	75	
3991	Pós-pago	Masculino		Gradiente	Nenhuma	Qualidade do sinal	3,657	5	64	75	
4038	Pré-pago	Feminino		Motorola	Nenhuma	Maior cobertura	2,841	6	72,8	76	
4748	Pré-pago	Feminino		Samsung	Nenhuma	Tarifa menor	3,849	8	94,6	86	

Figura 25 - Recodificação da variável Opinião - correspondência das células vazias

Agora podemos passar para a recodificação de variáveis quantitativas.

3.2 – Recodificação de Variáveis Quantitativas

A recodificação de variáveis quantitativas pode ter duas finalidades:

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

Em ambos os casos o objetivo é obter uma nova variável que facilite a análise dos dados, tornando-a mais resumida ou significativa. Por exemplo, vamos realizar duas recodificações da variável Renda, no arquivo Celular01.xls:

- transformá-la de Renda em salários mínimos para Renda em reais, simplesmente multiplicando seus valores pelo salário mínimo vigente (julho de 2010), 510 reais;
- recodificá-la em uma variável qualitativa, criando faixas de valores que definirão clientes de renda baixa, média e alta.

Tais ações facilitarão o estudo da Renda, posto que algumas pessoas têm dificuldade de raciocinar em termos de salários mínimos, e outras preferem uma visão “categorizada” das quantidades.

Tal como nas variáveis qualitativas, é preciso tomar cuidado com as células vazias, pois vamos “criar” informações ao codificá-la em qualitativa, ou causar erro ao transformá-la. Precisamos modificar a transformação, incluindo uma função SE, para que a multiplicação por 510 seja feita apenas quando as células de Renda NÃO forem vazias, para que não haja erros e tampouco “criação de informação”. De forma análoga à utilizada no caso da variável Opinião:

SE Renda = “ “ ENTÃO RendaR\$ = “ “ SENÃO RendaR\$ = Renda * 510

Ou, em termos da sintaxe do Excel: =SE(G2="";G2;G2*510)

A implementação no Excel pode ser vista na Figura 26.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69	Negativa	=SE(G2="" ;G2;G2*460)	
3	Pré-pago	Feminino	Mediana	Ericsson	Cu	Maior cobertura	1,491	4	50,2	66	Negativa	685,86	

Figura 26 - Função SE para transformar variável Renda - considerando células vazias

Arrastando a fórmula até a célula L5001 a variável estará corretamente transformada. Se repetirmos a filtragem na variável RendaR\$, escolhendo a exibição das células vazias, o resultado será a Figura 27.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$
1109	Pré-pago	Feminino	Mediana	Ericsson	Ferocom	Maior cobertura		5	56,6	60	Negativa	
1181	Pós-pago	Feminino	Sofrível	Samsung	Escuridão	Maior cobertura		2	33	79	Negativa	
3293	Pós-pago	Masculino	Terrível	Motorola	Mortinho	Tarifa menor		3	40,6	97	Negativa	
4027	Pós-pago	Masculino	Sofrível	Gradiente	Mortinho	Qualidade do sinal		7	81,8	73	Negativa	

Figura 27 - Transformação da variável Renda - correspondência das células vazias

Podemos agora pensar na recodificação de Renda em uma variável qualitativa, que chamaremos RendaC, que ocupará a coluna M. Vamos criar três classes *arbitrárias* de renda: os clientes com Renda de até 4 salários mínimos serão chamados RendaC Baixa, os com Renda entre 4 e 12 serão chamados RendaC Média, e os com Renda acima de 12 salários mínimos serão RendaC Alta. Trata-se de um caso que exige a utilização de funções SE: precisaremos de duas para a classificação da Renda, e mais uma para lidar com as células vazias. Veja a estrutura na Figura 28.

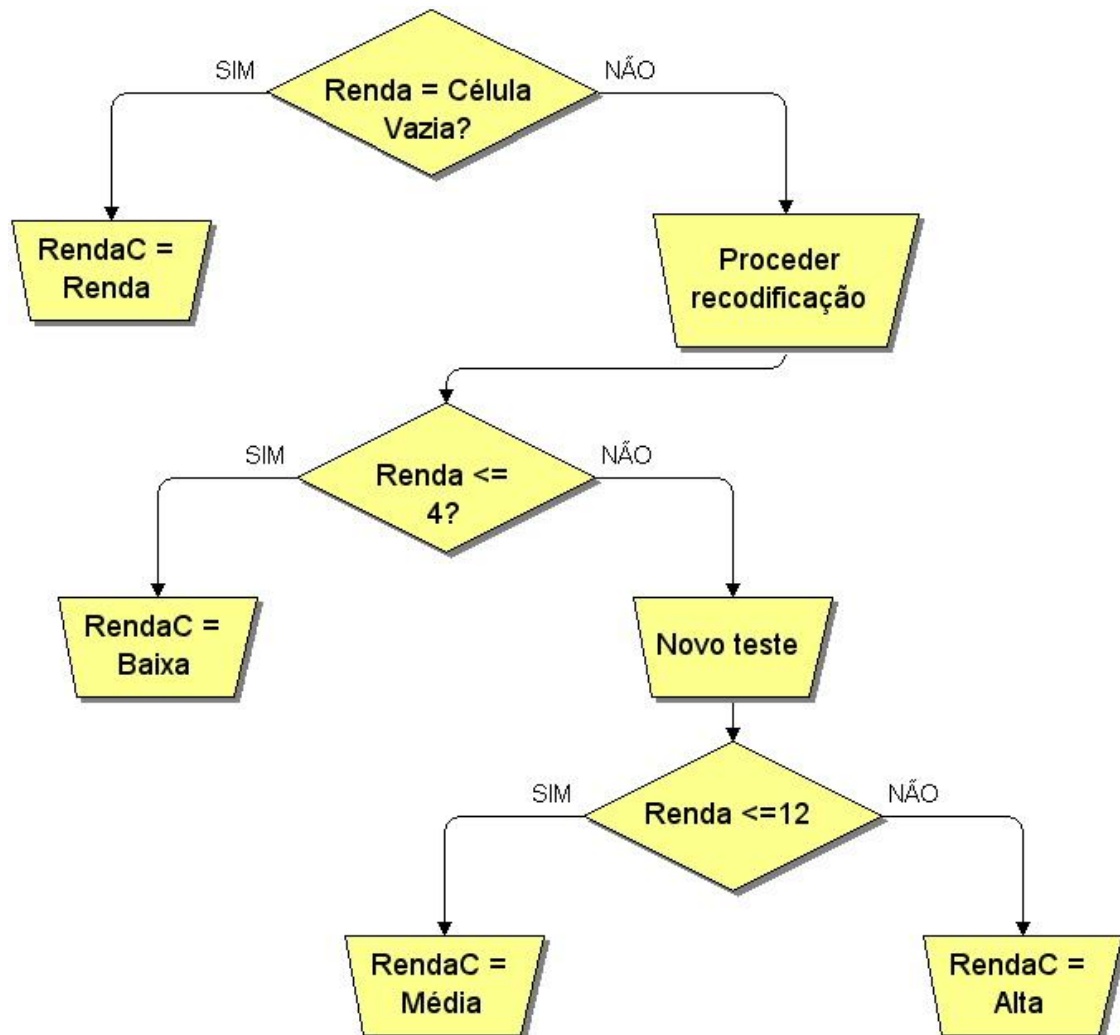


Figura 28 - Estrutura de recodificação da variável Renda

Vamos ver os passos:

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

Observe a implementação da recodificação da variável Renda no Excel (Figura 29).

	G	H	I	J	K	L	M	N	O	P	Q	R
1	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	RendaC					
2	1,074	3	39	69	Negativa	494,04	=SE(G2=" ";G2;SE(G2<=4;"Baixa";SE(G2<=12;"Média";"Alta")))					
3	1,491	4	50,2	66	Negativa	685,86						
4	2,778	7	80,8	68	Negativa	1277,88						

Figura 29 - Função de recodificação da variável Renda em RendaC

Ao arrastar a fórmula até a célula M5001 completamos a recodificação da variável. Se aplicarmos filtragem à RendaC observaremos apenas os valores “esperados”: células vazias, Baixa, Média e Alta, como na Figura 30.

	G	H	I	J	K	L	M
1	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	RendaC
2	1,074	3	39	69	Negativa	494	(Tudo)
3	1,491	4	50,2	66	Negativa	685	(10 Primeiros...)
4	2,778	7	80,8	68	Negativa	1277	(Personalizar...)
5	4,953	9	100,2	68	Negativa	2278	Alta
6	2,748	3	43,6	71	Negativa	1264	Baixa
7	1,924	5	59	65	Positiva	885	Média
8	1,773	4	49		Positiva	815,58	(Vazias)
9	6,552	5	63,8	78	Negativa	3013,92	(NãoVazias)

Figura 30 - Filtro da variável recodificada RendaC

Podemos fazer a recodificação envolvendo mais de uma variável. Por exemplo, imagine que a operadora ATCHIM resolveu focar seus esforços promocionais nos clientes de baixa renda (tal como definidos em RendaC, até 4 salários mínimos) e que têm opiniões positivas sobre a empresa (Confiável e Excelente, definidas como Positiva em OpiniãoC). Tal cliente é considerado como “Cliente alvo” das ofertas de planos especiais, novos aparelhos e serviços. Podemos criar uma nova variável, ClienteAlvo, recodificando RendaC e OpiniãoC.

Trata-se de mais um caso para uso da função SE, lembrando que devemos levar em conta as células vazias, mas com mais cuidado ainda: RendaC pode ter células vazias para alguns clientes, OpiniãoC também, ou, eventualmente ambas podem estar vazias para o mesmo cliente. Isso causará um impasse, pois não podemos criar recodificá-las em ClienteAlvo para as linhas em que pelo menos uma das variáveis apresentar células vazias. A solução é usar uma função OU: se OpiniãoC ou RendaC apresentarem células vazias, o valor correspondente de ClienteAlvo ficará vazia também. Se OpiniãoC for “Positiva” e RendaC for “Baixa”, ClienteAlvo será “Sim”, senão será “Não”. De forma análoga à utilizada no caso das variáveis OpiniãoC e RendaC:

SE OpiniãoC = “ “ OU RendaC = “ “ ENTÃO ClienteAlvo = “ “
 SENÃO SE OpiniãoC = “Positiva” E RendaC = “Baixa” ENTÃO ClienteAlvo = “Sim”
 SENÃO ClienteAlvo = “Não”

Ou, em termos da sintaxe do Excel:

=SE(OU(K2=" ";M2=" ");" ";SE(E(K2="Positiva";M2="Baixa");"Sim";"Não"))

A implementação no Excel pode ser vista na Figura 31.

	K	L	M	N	O	P	Q	R	S	T
1	OpiniãoC	RendaR\$	RendaC	ClienteAlvo						
2	Negativa	494,04	Baixa	=SE(OU(K2=" ";M2=" ");" ";SE(E(K2="Positiva";M2="Baixa");"Sim";"Não"))						
3	Negativa	685,86	Baixa							

Figura 31 - Variável ClienteAlvo: resultado da recodificação de OpiniãoC e RendaC

Arrastando a fórmula até a célula N5001 completa-se a criação da variável ClienteAlvo. Se aplicarmos filtragem à ClienteAlvo observaremos apenas os valores “esperados”: células vazias, Não e Sim, como na Figura 32.

	K	L	M	N
1	OpiniãoC	RendaR\$	RendaC	ClienteAlvo
2	Negativa	494,04	Baixa	(Tudo)
3	Negativa	685,86	Baixa	(10 Primeiros...)
4	Negativa	1277,88	Baixa	(Personalizar...)
5	Negativa	2278,38	Média	Não
6	Negativa	1264,08	Baixa	Sim
7	Positiva	885,04	Baixa	(Vazias)
8	Positiva	815,58	Baixa	(Não Vazias)
9	Negativa	3013,92	Média	Sim
				Não

Figura 32 - Filtro da variável ClienteAlvo

Com as operações de filtragem, classificação recodificação e transformação podemos identificar várias características interessantes dos dados, bem como agrupar valores das variáveis, o que facilitará a análise posterior das variáveis.

4. Análise de dados de variáveis qualitativas – com tabelas dinâmicas

Uma solução interessante para a análise de dados é o uso de tabelas dinâmicas, ferramenta incorporada ao Microsoft Excel que pode ser muito útil para extrair informações de um arquivo de dados. O uso da tabela dinâmica exige que os dados estejam com a formatação usual de um banco de dados: variáveis nas colunas e dados nas linhas, sendo que a primeira linha do arquivo deve conter os nomes das variáveis.

4.1 – Uma variável qualitativa

Imagine que estamos interessados nas variáveis qualitativas do arquivo Celular1.xls, por exemplo, a variável Tipo, já estudada anteriormente. Para construir uma tabela dinâmica devemos procurar no menu “Dados” pela opção “Relatório de tabela e gráfico dinâmicos”, ver Figura 33.

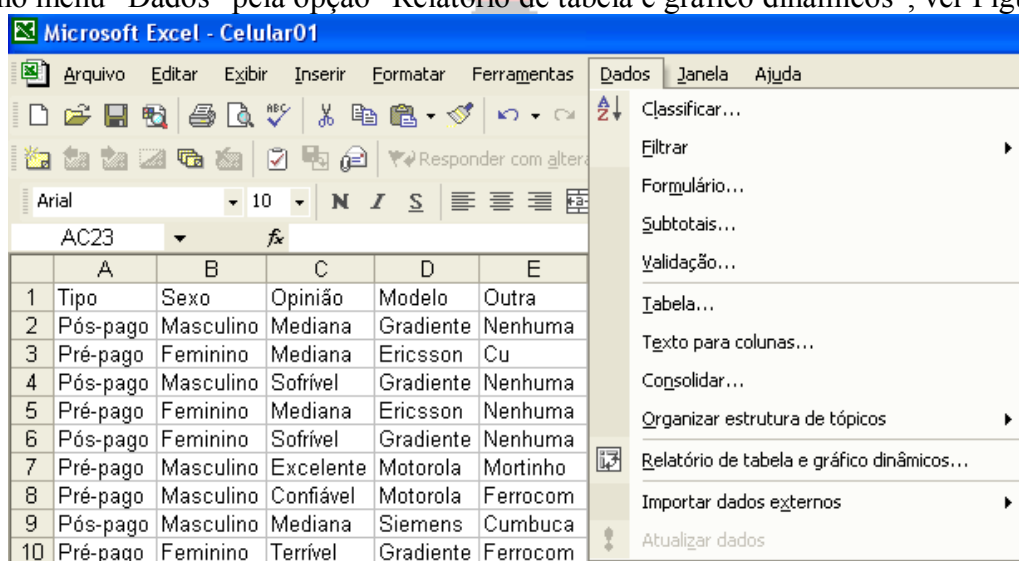


Figura 33 - Opções do Menu Dados

Ao escolher esta opção surgirá uma tela parecida com a Figura 34.

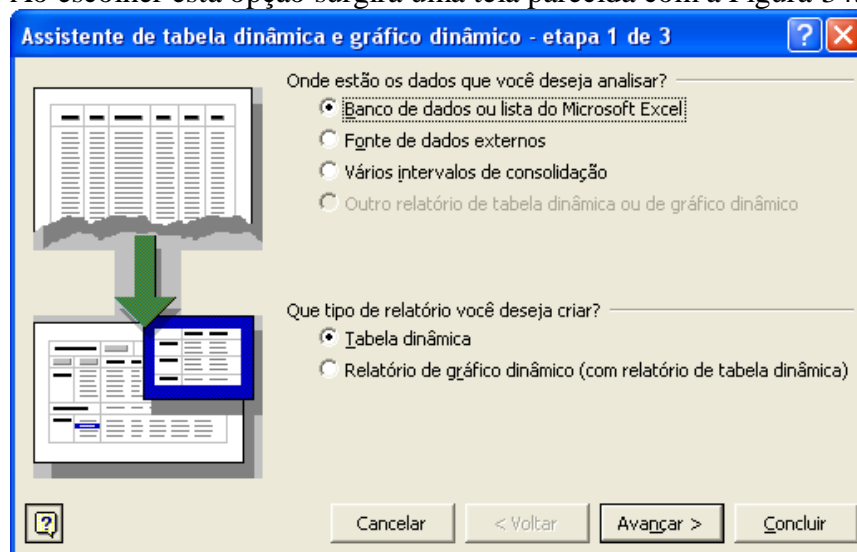


Figura 34 - Etapa 1 da Tabela Dinâmica

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 35.

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). Ver Figura 35.

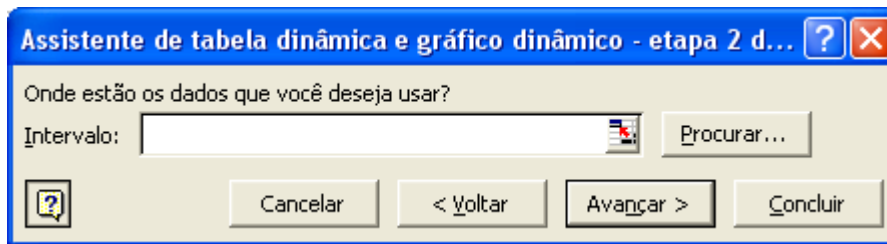


Figura 35 - Etapa 2 da Tabela Dinâmica

Muitas vezes, o Excel automaticamente seleciona todos os dados disponíveis na planilha (no caso da célula A1 até a célula N5001, compreendendo as 5000 observações de cada uma das 10 variáveis). Contudo, se isso não ocorrer, podemos selecionar o intervalo por conta própria, de duas maneiras:

- digitando no campo "Intervalo" (ver Figura 36) as referências das células (A1:N5001, 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, e selecionar as células na planilha, o que resultará na Figura 89 também.

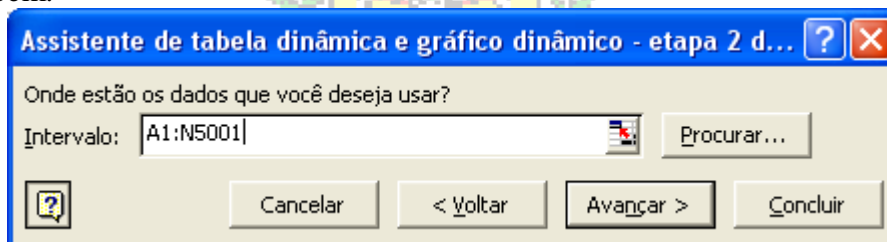
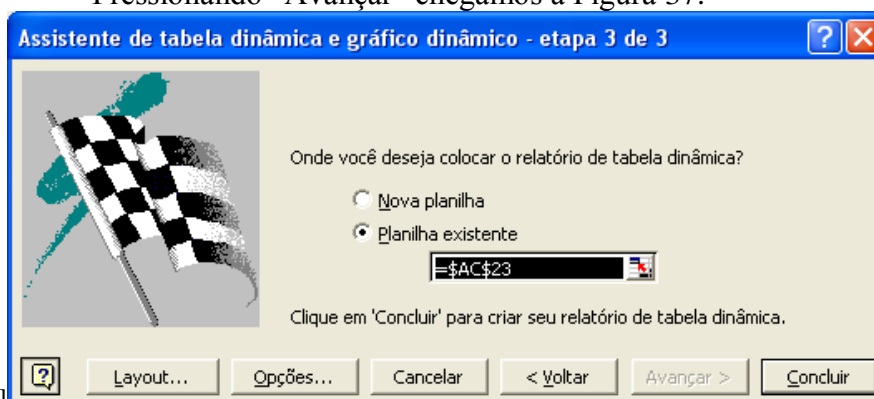


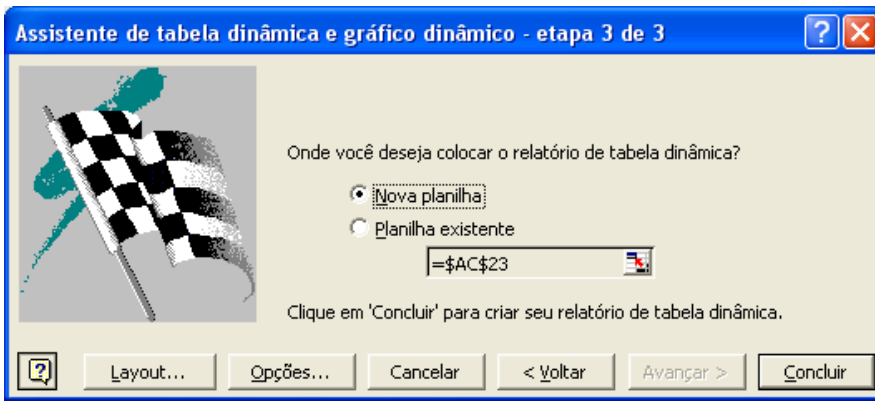
Figura 36 - Etapa 2 da Tabela Dinâmica – Dados selecionados

Pressionando “Avançar” chegamos à Figura 37.



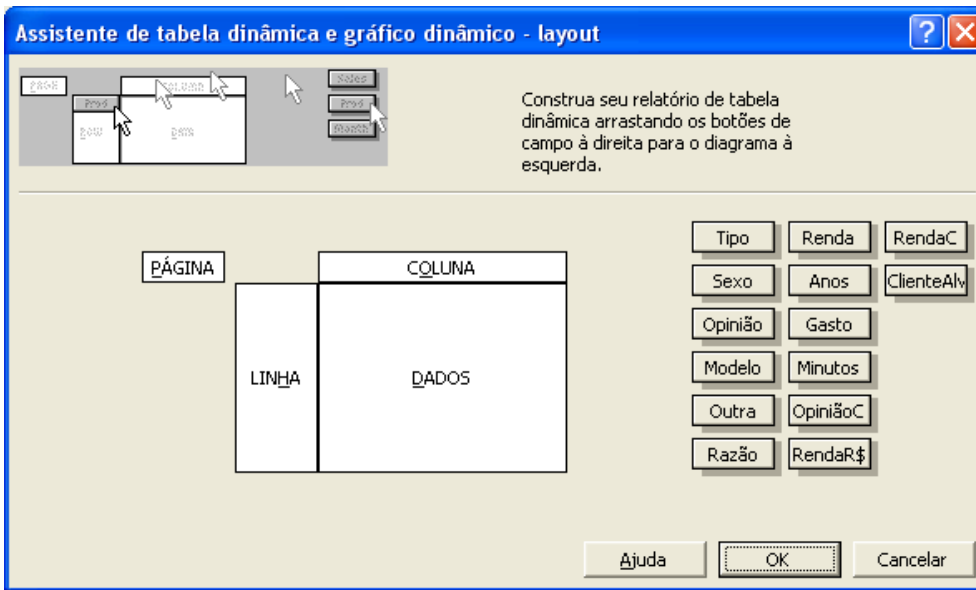
Podemos escolher onde os resultados da tabela serão apresentados: em uma célula específica de uma planilha existente (Figura 37) ou em uma nova planilha criada pelo Excel (Figura 38).

Figura 37 - Etapa 3 da Tabela Dinâmica – planilha existente



Até agora nada fizemos além de selecionar os dados de entrada e onde serão postos os resultados. É preciso definir o layout da tabela, pressionando a opção "Layout" na Figura 91, chegamos à Figura 39.

Figura 38 - Etapa 3 da Tabela Dinâmica - nova planilha



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 Tipo, devemos selecioná-la e arrastá-la até o campo "LINHA", ou "COLUNA". As outras variáveis não farão parte da tabela.

Figura 39 - Layout da tabela dinâmica

Devemos também arrastar a variável Tipo 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 (para as quantitativas é a soma), tal como mostrado na figura 40.

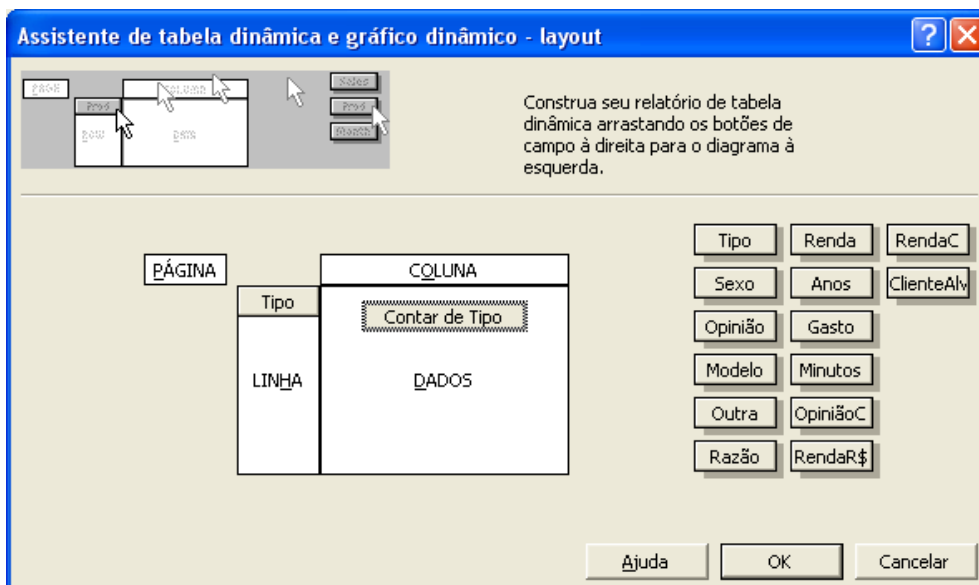


Figura 40 - Layout da tabela dinâmica para Tipo

A forma como o Excel apresentará a contagem de Tipo pode ser mudada facilmente, clicando duas vezes sobre o botão “Contar de Tipo” no campo Dados. Veja a Figura 41:

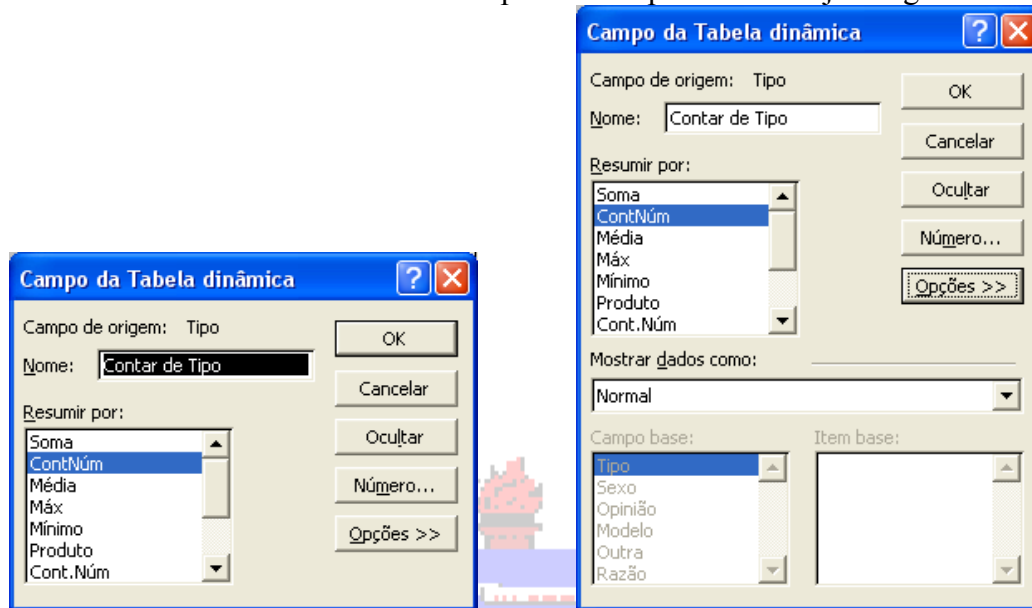


Figura 41 - Opções de Campo da Tabela Dinâmica

Inicialmente surge a tela da esquerda. Repare que podemos escolher resumir por: média, mínimo, máximo, desvio padrão, que serão úteis quando formos analisar variáveis quantitativas. Se pressionarmos “Opções”, a tela expandida à direita será mostrada pelo Excel. E podemos mostrar dados como: normal (simples contagem), ou como percentual em relação a uma referência, etc. Na tabela dinâmica podemos acrescentar mais um campo de dados com a variável Tipo, como mostrado na Figura 42, arrastando mais uma vez a variável Tipo para o campo Dados.



Figura 42 - Layout da tabela dinâmica de Tipo: dois campos

Da forma como está teremos uma tabela com duas colunas para frequências, o que não é o nosso interesse. Devemos mudar a forma como os dados são mostrados no campo “Contar de Tipo2”, de Normal (frequências) para % do total. Para tanto, basta clicar duas vezes sobre “Contar de Tipo2”, surgirá a tela à esquerda da Figura 41. Pressionando “Opções” chegamos à tela da direita da Figura 41. Selecionando a seta ao lado de normal, chegamos à Figura 43.

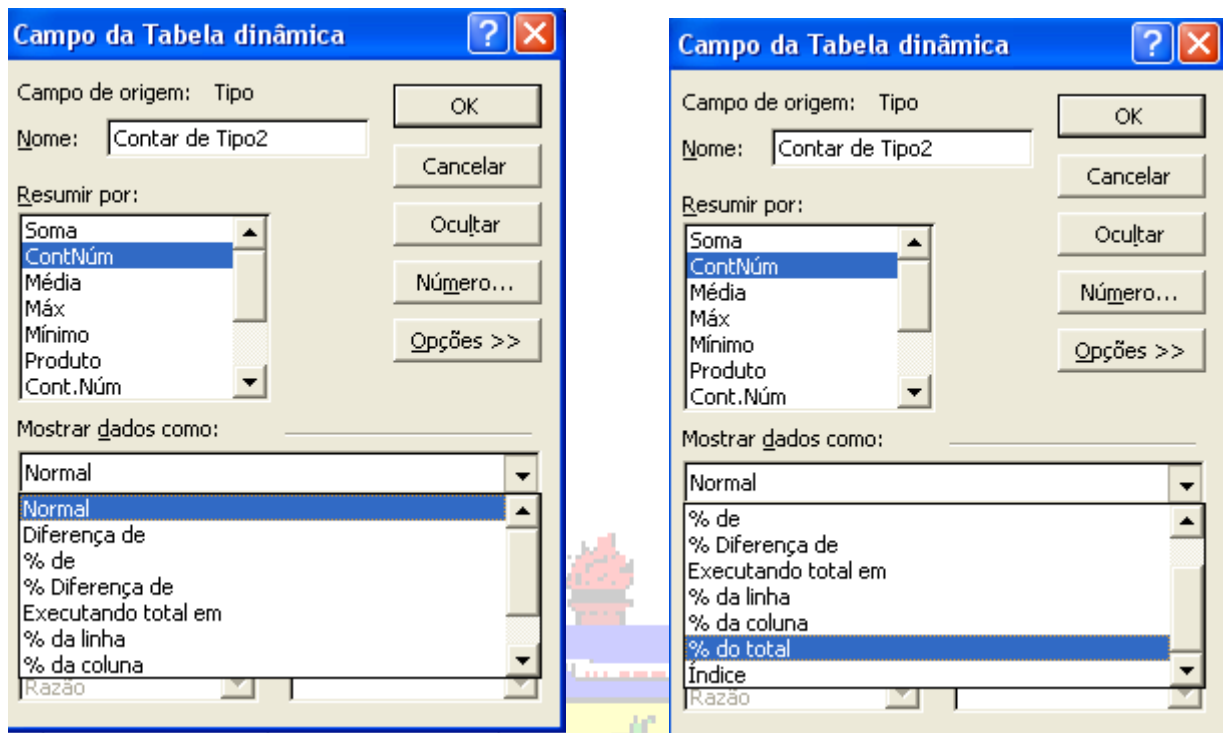


Figura 43 - Mudando a forma de mostrar os dados

Na tela à esquerda vemos as várias formas de mostrar os dados: sendo Normal a padrão (os dados serão mostrados no formato especificado no campo “Resumir por”), neste caso frequência (contagem). Mas, queremos mudar de frequência para percentual, em relação ao total dos dados: escolhemos então % do total, na tela à direita. Este procedimento resulta na Figura 44.

Microsoft Excel - Celular01

Arquivo Editar Exibir Inserir Formatar Ferramentas Dados Janela Ajuda

Arial 10

	A	B	C	D	E	F	G
1	Solte campos de página aqui						
2							
3	Tipo	Dados	Total				
4		Contar de Tipo	7				
5		Contar de Tipo2	0,14%				
6	Pós-pago	Contar de Tipo	2102				
7		Contar de Tipo2	42,04%				
8	Pré-pago	Contar de Tipo	2891				
9		Contar de Tipo2	57,82%				
10	Total	Contar de Tipo	5000				
11	Total	Contar de Tipo2	100,00%				
12							
13							
14							
15							
16							
17							
18							
19							
20							
21							
22							

Lista de campos da tabela dinâmica

Arraste itens para o relatório de tabela dinâmica

- Tipo
- Sexo
- Opinião
- Modelo
- Outra
- Razão
- Renda
- Anos
- Gasto
- Minutos

Adicionar a Área de linha

Tabela dinâmica

Tabela dinâmica

Figura 44 - Tabela dinâmica de Tipo 1a parte

Observe que todos os valores possíveis de Tipo são apresentados, automaticamente: Pós-pago, Pré-pago e células vazias. Conforme foi estabelecido no layout, para cada valor foi contada a frequência e calculado o respectivo percentual em relação ao total geral (5000).

É possível ver também uma caixa com opções (Lista de campos da tabela dinâmica) – que serve para se quisermos inserir mais variáveis na tabela – e uma barra de ferramentas (Tabela dinâmica) – que serve para formatar a tabela, atualizar dados, aumentar ou diminuir o detalhamento, ocultar campos, etc.

Se fecharmos ambos, o resultado será a distribuição de frequências da variável Tipo, conforme mostrado na Figura 45.

	A	B	C
1			
2			
3	Tipo	Dados	Total
4		Contar de Tipo	7
5		Contar de Tipo2	0,14%
6	Pós-pago	Contar de Tipo	2102
7		Contar de Tipo2	42,04%
8	Pré-pago	Contar de Tipo	2891
9		Contar de Tipo2	57,82%
10	Total Contar de Tipo		5000
11	Total Contar de Tipo2		100,00%

Figura 45 - Distribuição de freqüências de Tipo

	A	B	C
1			
2			
3	Tipo	Dados	Total
4			7
5			,14%
6			2102
7			,04%
8			2891
9			,82%
10			5000
11			,00%
12			
13			
14			
15			
16			

Figura 46 - Opções de apresentação de Tipo

Podemos observar que os clientes preferem o serviço Pré-pago, pois 2891 pessoas, 57,82% do total, escolheram este tipo. Já o Pós-pago foi escolhido por 42,04% dos clientes.

É importante notar também as células vazias: há 7 delas, o que corresponde a 0,14% do total de 5000 clientes, número aceitável.

As setas ao lado de Tipo e Dados permitem escolher o que queremos mostrar na tabela. Por exemplo, imagine que quiséssemos remover a linha com as células vazias. Veja a Figura 46.



Ao pressionarmos a seta ao lado de Tipo surge a tela da Figura 46. O padrão é “Mostrar tudo”: todos os valores da variável em questão, uma vez detectados pelo Excel, serão apresentados na tabela, com as características definidas na etapa de Layout (contagens, percentuais, etc.). Isso inclui as células vazias.

Pode ser de nosso interesse remover as células vazias da tabela. Se assim decidirmos, precisamos desmarcar a respectiva opção, tal como mostrado na tela à esquerda da Figura 47. O resultado final está na tela à direita da Figura 47, onde podemos perceber que os percentuais foram recalculados, agora tomando como referência o total de células não vazias, 4993. As conclusões serão semelhantes, apenas com uma ligeira modificação nos valores dos percentuais.

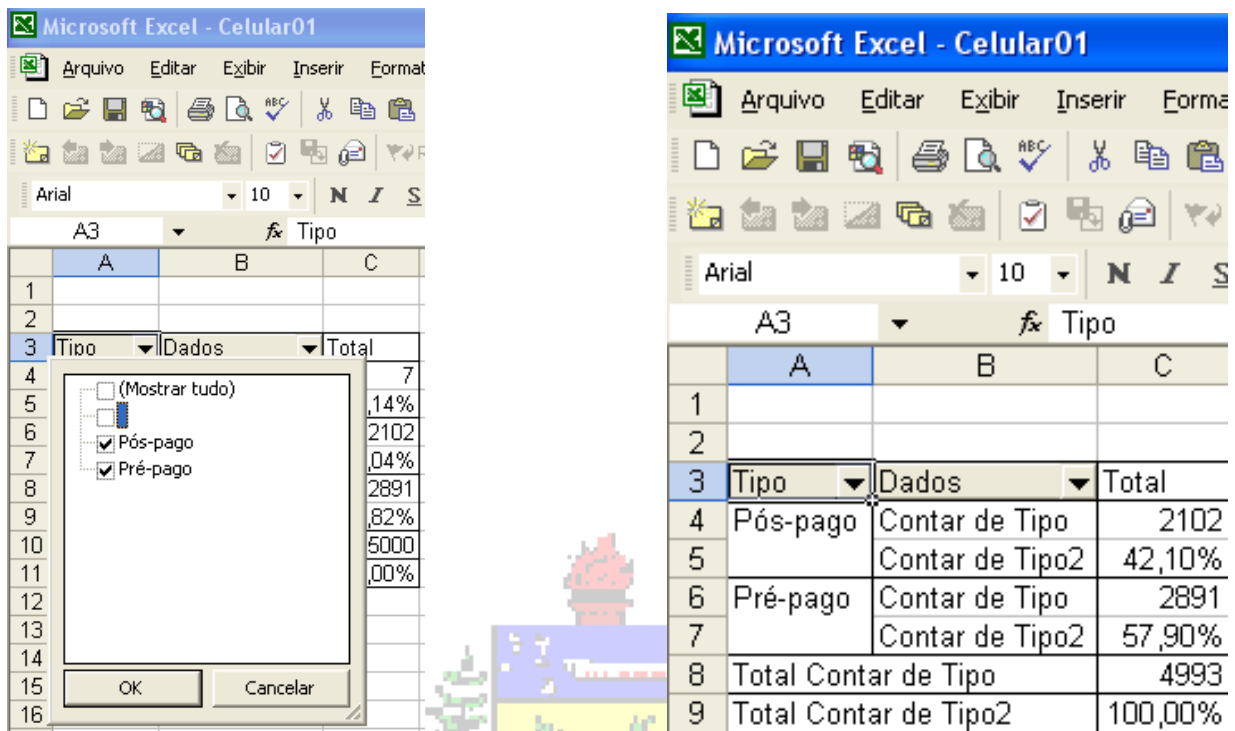


Figura 47 - Tabela dinâmica de Tipo: remoção de células vazias

Observa-se que a tabela apresenta algumas “poluições”, o que pode confundir sua interpretação por leigos: “Contar de Tipo”, “Contar de Tipo2”. Todos estes podem ser modificados, simplesmente por digitação dos nomes de nosso interesse, resultando na Figura 48.

Tipo	Dados	Total
Pós-pago	Frequência	2102
	%	42,10%
Pré-pago	Frequência	2891
	%	57,90%
Total Frequência		4993
Total %		100,00%

Figura 48 - Tabela dinâmica de Tipo: frequências absolutas e relativas

4.2 – 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 4.1: menu Dados, opção Relatório de Tabela e Gráfico Dinâmicos. Posteriormente, precisamos definir onde estão os dados: no nosso caso, na Plan1, células A1 a N5001 do arquivo Celular01.xls, como mostrado na Figura 49.

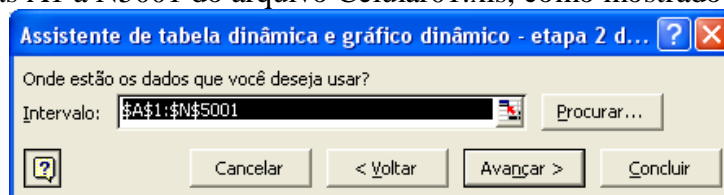


Figura 49 – Etapa 2 da Tabela Dinâmica

Como são os mesmos dados usados na tabela dinâmica da seção anterior o Excel apresentará a mensagem da Figura 50.

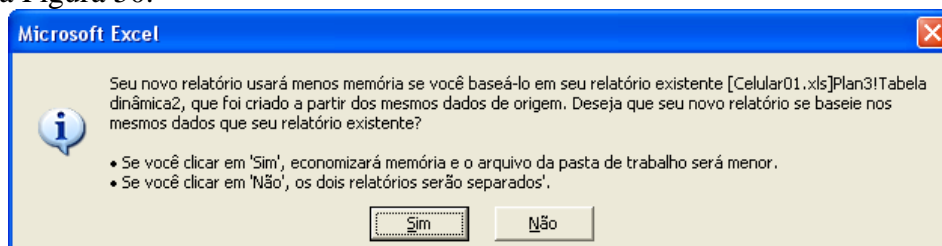


Figura 50 - Mensagem de alerta do Excel sobre tabela dinâmica existentes com os mesmos dados

Pressionando Sim na mensagem acima, economiza-se memória, levando à Figura 51.

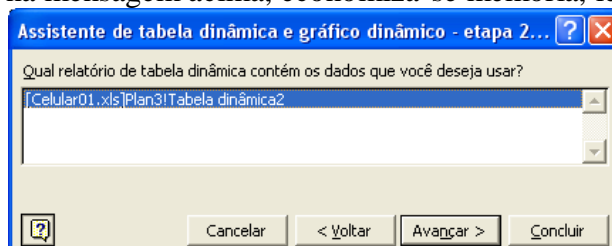


Figura 51 - Relatórios de tabela dinâmica disponíveis para os mesmos dados

No nosso caso há apenas uma tabela dinâmica disponível, e, claro, devemos escolhê-la e pressionar “Avançar”. Chegamos então à tela da Figura 52.

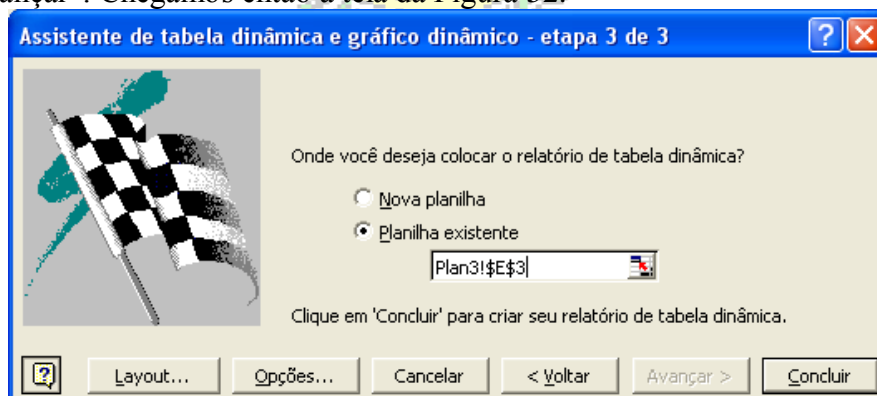


Figura 52 - Etapa 3 da tabela dinâmica

Desta vez vamos pôr os resultados em uma planilha existente, a mesma onde está a tabela dinâmica da variável Tipo, célula E3 (ou outra qualquer). Basta agora definir o Layout, pressionando o respectivo botão. Queremos fazer o cruzamento das variáveis Tipo e Opinião: selecione e arraste a variável Tipo para o campo "LINHA", e a variável Opinião para o campo "COLUNA" (ou vice-versa), como na Figura 53.

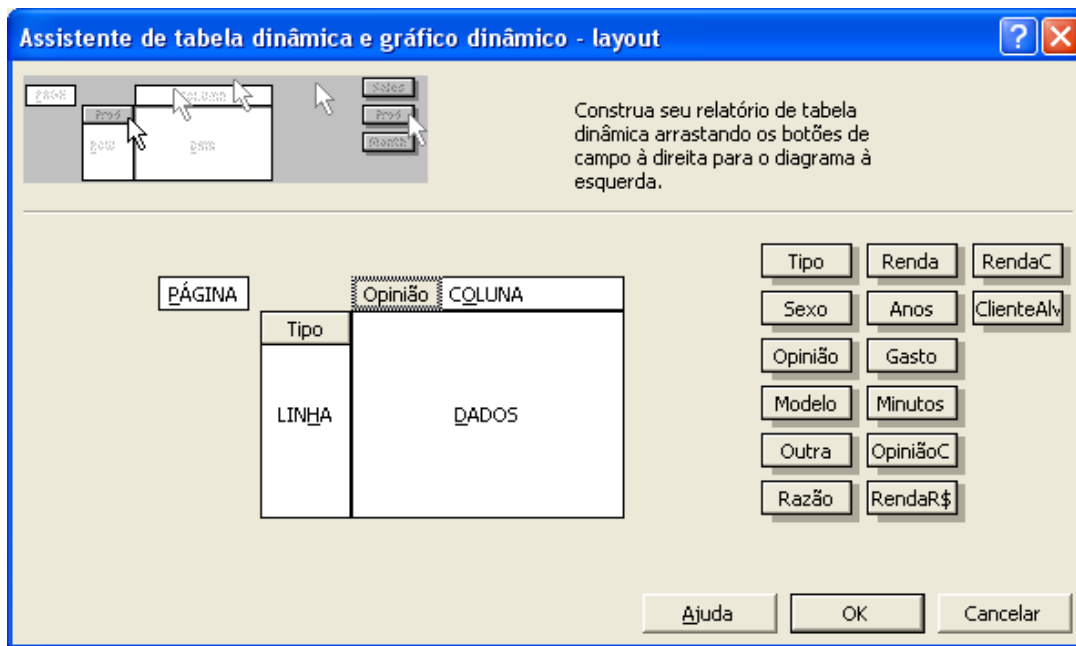


Figura 53 - Layout da tabela dinâmica Tipo x Opinião 1a parte

Precisamos agora preencher o campo Dados. Selecione a variável Opinião (ou a variável Tipo, mas APENAS uma delas) e arraste para o campo "DADOS": a ação padrão é "Contagem de Opinião" (ou "Contagem de Tipo", se você escolher esta variável), pois ambas as variáveis são qualitativas. Se as duas variáveis forem postas no campo "DADOS" o Excel construirá duas tabelas separadas, e NÃO cruzará as duas variáveis, que é o nosso objetivo. Veja a Figura 54.

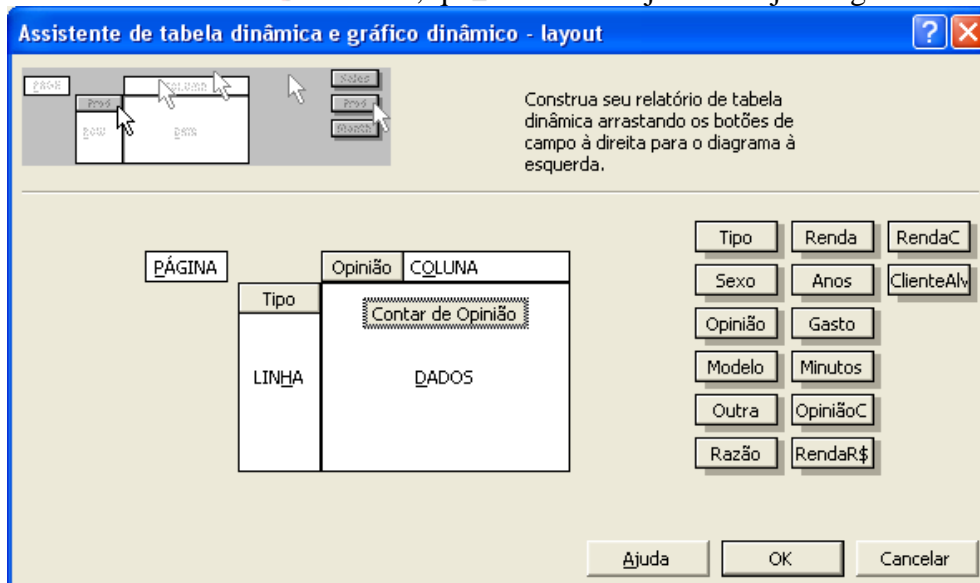


Figura 54 - Layout da tabela dinâmica Tipo x Opinião 2a parte

Ao construir uma tabela de contingências usualmente não temos interesse apenas nas frequências de cada cruzamento. Foi visto na seção 4.1 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 Opinião mais três vezes ao campo "DADOS", Figura 55.

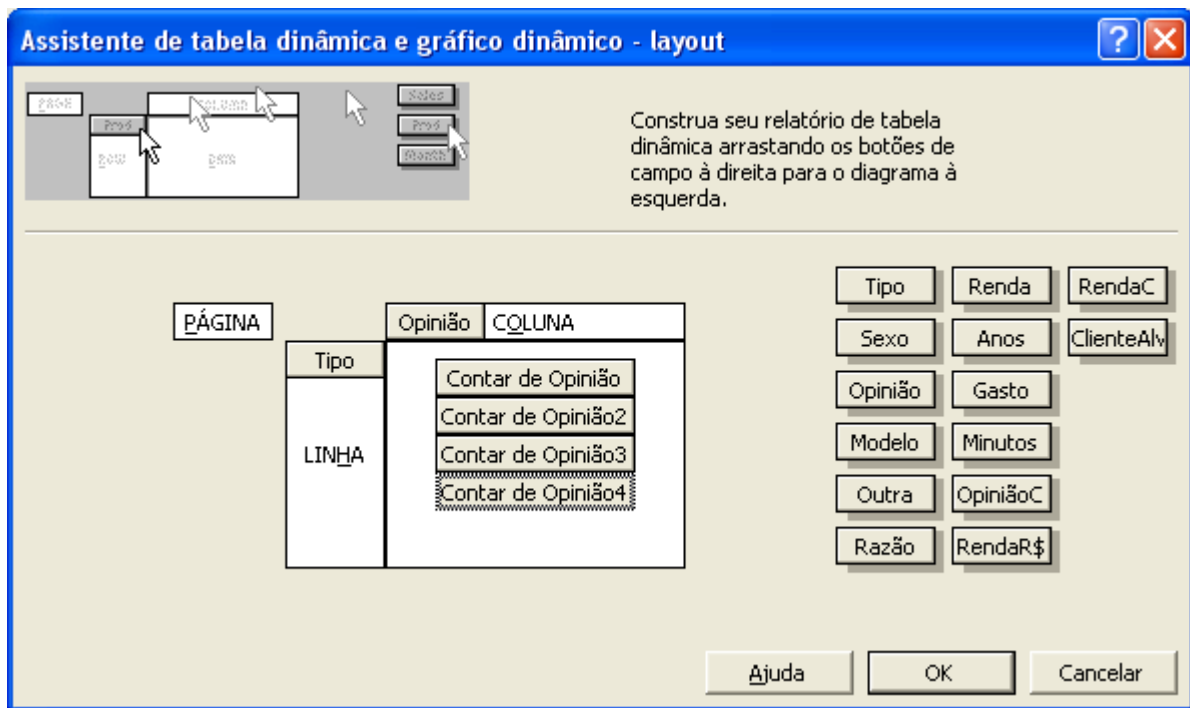


Figura 55 - Layout da tabela dinâmica Tipo x Opinião 3a parte

Precisamos modificar a forma como os dados dos botões Contar de Opinião2, Contar de Opinião3 e Contar de Opinião4 são mostrados, de Normal (frequências) para os percentuais apropriados. Começando com Contar de Opinião2, basta clicar duas vezes sobre o botão, resultando na Figura 56.

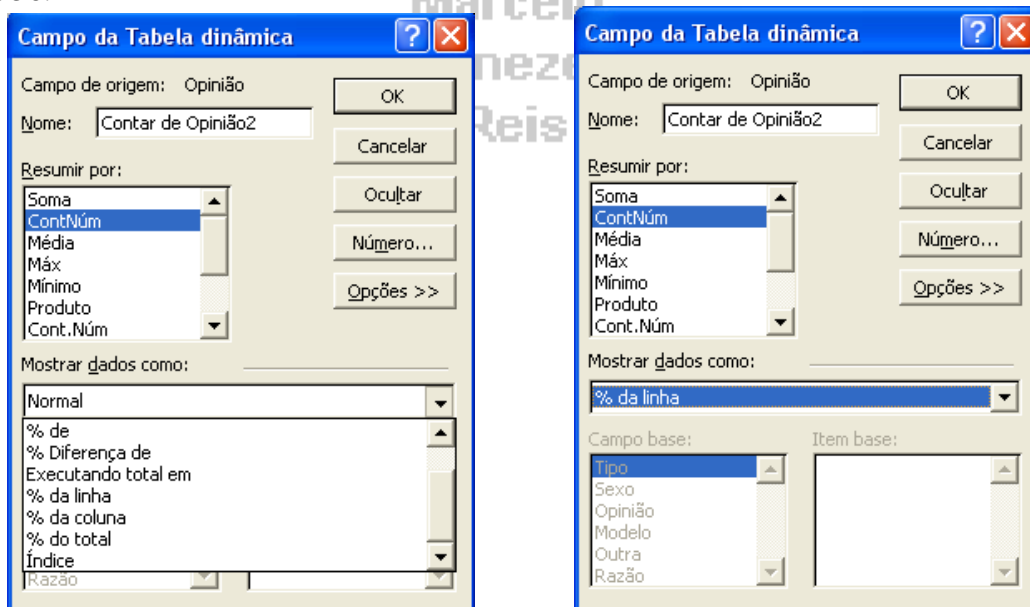


Figura 56 - Opções de percentuais por linhas

Para modificarmos a forma de mostrar os dados precisamos pressionar “Opções” (tela à esquerda na Figura 56), e escolher dentre as várias formas disponíveis a “% da linha” (tela à direita da Figura 56). Por um procedimento análogo podemos modificar Contar de Opinião3 (para % da coluna – tela à esquerda na Figura 57) e Contar de Opinião4 (para % do total – tela à direita na Figura 57).

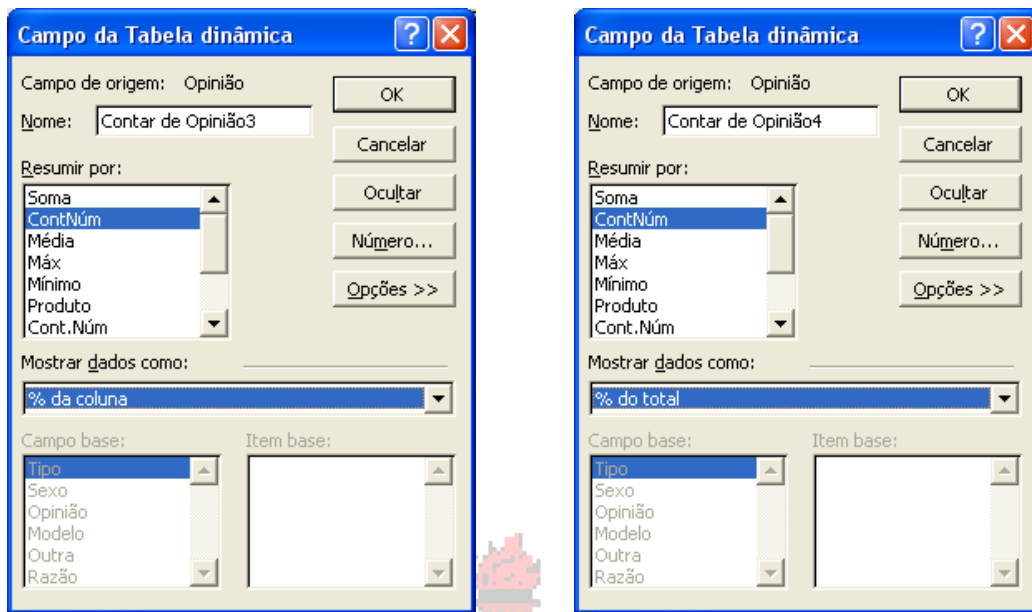


Figura 57 - Opções de percentuais por colunas e total

Finalmente a tabela está com o layout completo, resultando na Figura 58.

	D	E	F	G	H	I	J	K	L	M
1		Solte campos de página aqui								
2										
3				Opinião						
4		Tipo	Dados		Confiável	Excelente	Mediana	Sofrível	Terrível	Total geral
5			Contar de Opinião		1	1	2	2	1	7
6			Contar de Opinião2	0,00%	14,29%	14,29%	28,57%	28,57%	14,29%	100,00%
7			Contar de Opinião3	0,00%	0,13%	0				0,14%
8			Contar de Opinião4	0,00%	0,02%	0				0,14%
9		Pós-pago	Contar de Opinião		4	103				2102
10			Contar de Opinião2	0,19%	4,90%	6				100,00%
11			Contar de Opinião3	40,00%	13,41%	27				42,04%
12			Contar de Opinião4			2				42,04%
13			Total Contar de Opinião		10	768				2891
14			Contar de Opinião2	0,21%	22,97%	12				100,00%
15			Contar de Opinião3	60,00%	86,46%	72				57,82%
16			Contar de Opinião4	0,12%	13,28%	7				57,82%
17			Total Contar de Opinião		10	768				5000
18			Total Contar de Opinião2	0,20%	15,36%	10				100,00%
19			Total Contar de Opinião3	100,00%	100,00%	100				100,00%
20			Total Contar de Opinião4	0,20%	15,36%	10				100,00%
21										
22										
23										
24										

Figura 58 - Tabela dinâmica de Tipo x Opinião 1a parte

Da mesma forma que na tabela dinâmica de Tipo, podemos modificar vários aspectos da tabela dinâmica, ou, simplesmente fechar a Lista de campos e a barra de ferramentas, resultando na Figura 59.

	E	F	G	H	I	J	K	L	M
3			Opinião						
4	Tipo	Dados		Confiável	Excelente	Mediana	Sofrível	Terrível	Total geral
5		Contar de Opinião		1	1	2	2	1	7
6		Contar de Opinião2	0,00%	14,29%	14,29%	28,57%	28,57%	14,29%	100,00%
7		Contar de Opinião3	0,00%	0,13%	0,20%	0,14%	0,20%	0,08%	0,14%
8		Contar de Opinião4	0,00%	0,02%	0,02%	0,04%	0,04%	0,02%	0,14%
9	Pós-pago	Contar de Opinião	4	103	137	497	519	842	2102
10		Contar de Opinião2	0,19%	4,90%	6,52%	23,64%	24,69%	40,06%	100,00%
11		Contar de Opinião3	40,00%	13,41%	27,08%	34,02%	51,74%	67,25%	42,04%
12		Contar de Opinião4	0,08%	2,06%	2,74%	9,94%	10,38%	16,84%	42,04%
13	Pré-pago	Contar de Opinião	6	664	368	962	482	409	2891
14		Contar de Opinião2	0,21%	22,97%	12,73%	33,28%	16,67%	14,15%	100,00%
15		Contar de Opinião3	60,00%	86,46%	72,73%	65,85%	48,06%	32,67%	57,82%
16		Contar de Opinião4	0,12%	13,28%	7,36%	19,24%	9,64%	8,18%	57,82%
17	Total Contar de Opinião		10	768	506	1461	1003	1252	5000
18	Total Contar de Opinião2		0,20%	15,36%	10,12%	29,22%	20,06%	25,04%	100,00%
19	Total Contar de Opinião3		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
20	Total Contar de Opinião4		0,20%	15,36%	10,12%	29,22%	20,06%	25,04%	100,00%

Figura 59 - Tabela dinâmica de Tipo x Opinião 2a parte

Percebemos que na tabela há o registro de células vazias para Tipo e Opinião e os respectivos cruzamentos: não há nenhum caso de clientes com células vazias em ambas as variáveis. Mas, há 4 casos de Pós-pago e 6 de Pré-pago sem Opinião. E, 1 caso de Confiável, 1 de Excelente e 1 de Terrível sem Tipo de serviço, bem como 2 de Mediana e 2 de Sofrível na mesma situação. Estas células vazias, somadas das duas variáveis, representam 0,34% do total de 5000 observações, uma quantidade muito pequena, e que está “poluindo” a tabela.

Da mesma forma que fizemos na seção 4.1, para Tipo, podemos remover as células vazias da tabela dinâmica, desmarcando-as das listas de valores de Tipo e de Opinião, como mostrado nas Figuras 60 e 61.

	E	F	G	H	I	J	K	L	M
3			Opinião						
4	Tipo	Dados		Confiável	Excelente	Mediana	Sofrível	Terrível	Total geral
5		Contar de Opinião		1	1	2	2	1	7
6		Contar de Opinião2	0,00%	14,29%	14,29%	28,57%	28,57%	14,29%	100,00%
7		Contar de Opinião3	0,00%	0,13%	0,20%	0,14%	0,20%	0,08%	0,14%
8		Contar de Opinião4	0,00%	0,02%	0,02%	0,04%	0,04%	0,02%	0,14%
9		Contar de Opinião	4	103	137	497	519	842	2102
10		Contar de Opinião2	0,19%	4,90%	6,52%	23,64%	24,69%	40,06%	100,00%
11		Contar de Opinião3	40,00%	13,41%	27,08%	34,02%	51,74%	67,25%	42,04%
12		Contar de Opinião4	0,08%	2,06%	2,74%	9,94%	10,38%	16,84%	42,04%
13		Contar de Opinião	6	664	368	962	482	409	2891
14		Contar de Opinião2	0,21%	22,97%	12,73%	33,28%	16,67%	14,15%	100,00%
15		Contar de Opinião3	60,00%	86,46%	72,73%	65,85%	48,06%	32,67%	57,82%
16		Contar de Opinião4	0,12%	13,28%	7,36%	19,24%	9,64%	8,18%	57,82%
17	Total Contar de Opinião		10	768	506	1461	1003	1252	5000
18	Total Contar de Opinião2		0,20%	15,36%	10,12%	29,22%	20,06%	25,04%	100,00%
19	Total Contar de Opinião3		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
20	Total Contar de Opinião4		0,20%	15,36%	10,12%	29,22%	20,06%	25,04%	100,00%

Figura 60 - Tabela dinâmica de Tipo x Opinião 3a parte

	E	F	G	H	I	J	K	L	M
3			Opinião						
4	Tipo	Dados							
5	Pós-pago	Contar de Opinião				497	519	842	2102
6		Contar de Opinião2				23,64%	24,69%	40,06%	100,00%
7		Contar de Opinião3				34,06%	51,85%	67,31%	42,10%
8		Contar de Opinião4				9,95%	10,39%	16,86%	42,10%
9	Pré-pago	Contar de Opinião				962	482	409	2891
10		Contar de Opinião2				33,28%	16,67%	14,15%	100,00%
11		Contar de Opinião3				65,94%	48,15%	32,69%	57,90%
12		Contar de Opinião4				19,27%	9,65%	8,19%	57,90%
13	Total Contar de Opinião					1459	1001	1251	4993
14	Total Contar de Opinião2					29,22%	20,05%	25,06%	100,00%
15	Total Contar de Opinião3					100,00%	100,00%	100,00%	100,00%
16	Total Contar de Opinião4					29,22%	20,05%	25,06%	100,00%

Figura 61 - Tabela dinâmica de Tipo x Opinião 4a parte

E finalmente a tabela sem células vazias, com os percentuais recalculados, na Figura 62.

	E	F	G	H	I	J	K	L
3			Opinião					
4	Tipo	Dados	Confiável	Excelente	Mediana	Sofrível	Terrível	Total geral
5	Pós-pago	Contar de Opinião	103	137	497	519	842	2098
6		Contar de Opinião2	4,91%	6,53%	23,69%	24,74%	40,13%	100,00%
7		Contar de Opinião3	13,43%	27,13%	34,06%	51,85%	67,31%	42,10%
8		Contar de Opinião4	2,07%	2,75%	9,97%	10,42%	16,90%	42,10%
9	Pré-pago	Contar de Opinião	664	368	962	482	409	2885
10		Contar de Opinião2	23,02%	12,76%	33,34%	16,71%	14,18%	100,00%
11		Contar de Opinião3	86,57%	72,87%	65,94%	48,15%	32,69%	57,90%
12		Contar de Opinião4	13,33%	7,39%	19,31%	9,67%	8,21%	57,90%
13	Total Contar de Opinião		767	505	1459	1001	1251	4983
14	Total Contar de Opinião2		15,39%	10,13%	29,28%	20,09%	25,11%	100,00%
15	Total Contar de Opinião3		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
16	Total Contar de Opinião4		15,39%	10,13%	29,28%	20,09%	25,11%	100,00%

Figura 62 - Tabela dinâmica de Tipo x Opinião (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: Confiável, Excelente, Mediana, Sofrível, Terrível. 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 Opinião, mudando a posição de Confiável para a direita: basta selecionar Confiável, e pressionar o botão direito do mouse, resultando na Figura 63.

	E	F	G	H	I	J	K	L
3			Opinião					
4	Tipo	Dados	Confiável	Excelente	Mediana	Sofrível	Terrível	Total geral
5	Pós-pago	Contar de Opinião					842	2098
6		Contar de Opinião2					0,13%	100,00%
7		Contar de Opinião3					7,31%	42,10%
8		Contar de Opinião4					6,90%	42,10%
9	Pré-pago	Contar de Opinião					409	2885
10		Contar de Opinião2					4,18%	100,00%
11		Contar de Opinião3					2,69%	57,90%
12		Contar de Opinião4					8,21%	57,90%
13	Total Contar de Opinião						1251	4983
14	Total Contar de Opinião2						5,11%	100,00%
15	Total Contar de Opinião3						0,00%	100,00%
16	Total Contar de Opinião4						5,11%	100,00%

Figura 63 - Tabela dinâmica Tipo x Opinião - opções de formatação

Observe a opção Ordem na Figura 63, selecionando-a vamos à Figura 64.

	E	F	G	H	I	J	K	L	M
3			Opinião						
4	Tipo	Dados	Confiável	Excelente	Mediana	Sofrível	Terrível	Total geral	
5	Pós-pago	Contar de Opinião						842	2098
6		Contar de Opinião2						13%	100,00%
7		Contar de Opinião3						31%	42,10%
8		Contar de Opinião4						90%	42,10%
9	Pré-pago	Contar de Opinião						409	2885
10		Contar de Opinião2						18%	100,00%
11		Contar de Opinião3						69%	57,90%
12		Contar de Opinião4						21%	57,90%
13	Total Contar de Opinião							1251	4983
14	Total Contar de Opinião2							11%	100,00%
15	Total Contar de Opinião3								
16	Total Contar de Opinião4								

Figura 64 - Tabela dinâmica Tipo x Opinião - mudança de ordem de valores

Há duas opções possíveis: mover para a direita ou mover para o fim. Como precisamos mudar apenas Confiável escolhemos mover para a direita, resultando na Figura 65.

	E	F	G	H	I	J	K	L
3			Opinião					
4	Tipo	Dados	Excelente	Confiável	Mediana	Sofrível	Terrível	Total geral
5	Pós-pago	Contar de Opinião	137	103	497	519	842	2098
6		Contar de Opinião2	6,53%	4,91%	23,69%	24,74%	40,13%	100,00%
7		Contar de Opinião3	27,13%	13,43%	34,06%	51,85%	67,31%	42,10%
8		Contar de Opinião4	2,75%	2,07%	9,97%	10,42%	16,90%	42,10%
9	Pré-pago	Contar de Opinião	368	664	962	482	409	2885
10		Contar de Opinião2	12,76%	23,02%	33,34%	16,71%	14,18%	100,00%
11		Contar de Opinião3	72,87%	86,57%	65,94%	48,15%	32,69%	57,90%
12		Contar de Opinião4	7,39%	13,33%	19,31%	9,67%	8,21%	57,90%
13	Total Contar de Opinião		505	767	1459	1001	1251	4983
14	Total Contar de Opinião2		10,13%	15,39%	29,28%	20,09%	25,11%	100,00%
15	Total Contar de Opinião3		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
16	Total Contar de Opinião4		10,13%	15,39%	29,28%	20,09%	25,11%	100,00%

Figura 65 - Tabela dinâmica de Tipo x Opinião – 1ª versão

Observa-se que novamente a tabela apresenta algumas “poluições”, o que pode confundir sua interpretação por leigos: “Contar de Tipo”, “Contar de Tipo2”, etc. Todos estes podem ser modificados, simplesmente por digitação dos nomes de nosso interesse, resultando na Figura 66.

		Opinião					
Tipo	Dados	Excelente	Confiável	Mediana	Sofrível	Terrível	Total geral
Pós-pago	Frequências	137	103	497	519	842	2098
	% por linha	6,53%	4,91%	23,69%	24,74%	40,13%	100,00%
	% por coluna	27,13%	13,43%	34,06%	51,85%	67,31%	42,10%
	% do total	2,75%	2,07%	9,97%	10,42%	16,90%	42,10%
Pré-pago	Frequências	368	664	962	482	409	2885
	% por linha	12,76%	23,02%	33,34%	16,71%	14,18%	100,00%
	% por coluna	72,87%	86,57%	65,94%	48,15%	32,69%	57,90%
	% do total	7,39%	13,33%	19,31%	9,67%	8,21%	57,90%
Total Frequências		505	767	1459	1001	1251	4983
Total % por linha		10,13%	15,39%	29,28%	20,09%	25,11%	100,00%
Total % por coluna		100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Total % do total		10,13%	15,39%	29,28%	20,09%	25,11%	100,00%

Figura 66 - Tabela dinâmica de Tipo x Opinião – 2ª versão

Podemos comparar as opiniões em função do tipo de serviço. Sempre devemos comparar os percentuais dos cruzamentos (por LINHA ou por COLUNA) com percentuais em relação aos totais de Tipo ou Opinião.

No presente caso, imagine que quiséssemos avaliar se há diferença entre as opiniões dos clientes da operadora em função do tipo de serviço. Devemos, então, usar os percentuais por linha, comparando os resultados de pós-pago e pré-pago com o Total % por linha. Observe que: do total de clientes, 10,13% tem opinião Excelente, 15,39% Confiável, 29,28% Mediana, 20,09% Sofrível e 25,11% Terrível. Se houver diferença de opinião dependendo do tipo de serviço os percentuais por linha de cada tipo devem ser substancialmente diferentes (pelo menos 5%) desses valores. Na Figura 58 é possível constatar:

- 6,53% dos clientes de Pós-pago consideram a qualidade do serviço Excelente, contra 12,76% do Pré-pago e 10,13% no total geral (observe que o percentual do Pré-pago é quase o dobro do Pós-pago);
- 4,91% dos clientes de Pós-pago têm opinião Confiável, contra 23,02% do Pré-pago e 15,39% do Total (percentual do Pré-pago é mais de 4 vezes maior do que o Pós-pago, e ambos estão a mais de 5% do Total²);
- 23,69% dos clientes de Pós-pago têm opinião Mediana, contra 33,34% do Pré-pago e 29,28% do Total (percentuais do Pré-pago e Pós-pago estão a mais de 5% do Total);
- 24,74% dos clientes de Pós-pago têm opinião Sofrível, contra 16,71% do Pré-pago e 20,09% do Total (houve uma inversão em relação às opiniões anteriores, os clientes do Pós-pago parecem mais insatisfeitos);
- 40,13% dos clientes de Pós-pago têm opinião Terrível, contra apenas 14,18% do Pré-pago e 25,11% do Total (percentual do Pós-pago é mais de duas vezes maior do que o Pré-pago, e mais de 1,5 vezes maior do que o Total).

Conclui-se, portanto, que os clientes do Pós-pago estão mais insatisfeitos do que os do Pré-pago.

Além disso, temos os percentuais por coluna. Verifica-se que, do total de opiniões Excelente nada menos do que 72,87% são de clientes de Pré-pago, e 67,31% das opiniões Terrível são de Pós-pago (sendo que os de Pré-pago representam 57,90% do total de clientes e os de Pós-pago 42,10%³), mostrando claramente que estes últimos estão mais insatisfeitos do que aqueles,

² Se não houvesse relação entre Tipo e Opinião os percentuais de Pós-pago e Pré-pago deveriam ser próximos de 15,39%.

³ Novamente, se não houvesse relação entre Tipo e Opinião os percentuais de Excelente de Pós-pago deveriam ser próximos de 42,10% e de Pré-pago 57,90%; a mesma coisa para os percentuais de Terrível.

corroborando as conclusões da análise dos percentuais por linhas. A Figura 67 resume o comportamento dos percentuais e mostra como eles mudam em relação ao total geral dependendo do Tipo de serviço: as linhas de Pós-pago e Pré-pago cruzam a do Total geral, indicando mudança de opinião dependendo do Tipo.

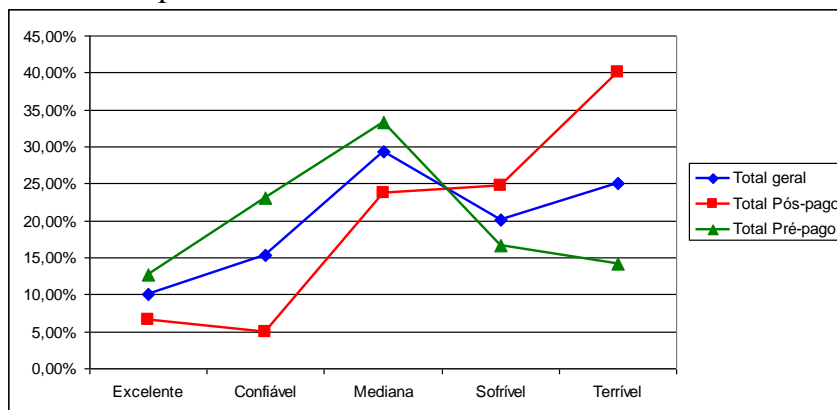


Figura 67⁴ - Percentuais de Opinião por Tipo de serviço

4.3 – Três Variáveis Qualitativas

Em tese poderíamos construir tabelas para cruzar qualquer número de variáveis qualitativas. Contudo, a interpretação dos resultados seria bastante difícil⁵. Mas, para os casos em que há no máximo 3 variáveis é plenamente possível construir e interpretar uma tabela dinâmica no Excel 2003. Vamos supor que queremos avaliar se homens ou mulheres, dependendo do tipo de serviço, estão mais satisfeitos (apresentam mais opiniões Excelente e Confiável). São três variáveis qualitativas: Sexo, Tipo e Opinião. Já construímos uma tabela dinâmica relacionando Tipo (nas linhas) e Opinião (nas colunas). O ideal seria obter uma tabela semelhante para o sexo feminino e outra para o masculino, de maneira a poder comparar os percentuais por linha, por exemplo. Isso pode ser feito com facilidade no Excel 2003 seguindo os passos usuais de montagem de tabela dinâmica, conforme visto na seção 4.2, para duas variáveis qualitativas.

- menu Dados, opção Relatório de Tabela e Gráfico Dinâmicos. Posteriormente, precisamos definir onde estão os dados: no nosso caso, na Plan1, células A1 a N5001 do arquivo Celular01.xls, como mostrado na Figura 49;

- seguem as instruções para economizar memória (Figuras 50 e 51);

- chega-se a uma tela parecida com a da Figura 52 (podemos colocar os resultados na mesma planilha onde estão as tabelas anteriores;

- basta agora definir o Layout, pressionando o respectivo botão (ver Figura 52);

#queremos fazer o cruzamento das variáveis Tipo e Opinião, em função de Sexo: selecione e arraste a variável Tipo para o campo "LINHA", e a variável Opinião para o campo "COLUNA" (ou vice-versa), tal como na seção 4.2;

#mas, a variável Sexo será arrastada para o campo "PÁGINA"⁶, o que permitirá usá-la como variável de controle⁷.

#para não sobrecarregar demais a tabela vamos inserir apenas duas medidas no campo "DADOS", frequência e percentuais por linha.

- o resultado final, já com as modificações nos nomes dos campos, Confiável movida para a direita, e as células vazias removidas, pode ser visto na Figura 68:

⁴ Esta figura é um gráfico de linhas, que estudaremos posteriormente.

⁵ Nestes casos é recomendável usar análise estatística multivariada.

⁶ No Excel 2007 este campo é chamado "Filtro de relatório".

⁷ Teoricamente podemos colocar várias variáveis neste campo, mas a interpretação dos dados pode tornar-se bem mais difícil.

Sexo		(Tudo)					
		Opinião					
Tipo	Dados	Excelente	Confiável	Mediana	Sofrível	Terrível	Total geral
Pós-pago	Frequências	137	103	497	519	842	2098
	% por linha	6,53%	4,91%	23,69%	24,74%	40,13%	100,00%
Pré-pago	Frequências	368	664	962	482	409	2885
	% por linha	12,76%	23,02%	33,34%	16,71%	14,18%	100,00%
Total Frequências		505	767	1459	1001	1251	4983
Total % por linha		10,13%	15,39%	29,28%	20,09%	25,11%	100,00%

Figura 68 - Tabela dinâmica de Tipo por Opinião em função do Sexo dos clientes

Trata-se de parte da tabela mostrada na Figura 66, porque na variável Sexo está selecionada a opção Tudo. Clicando sobre ▼ ao lado de Sexo, vemos as opções de exibição da Figura 69.

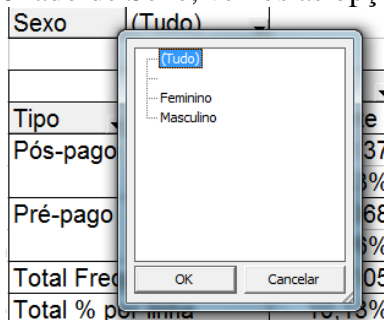


Figura 69 - Opções de exibição da variável Sexo

As possibilidades são: células vazias (-), Feminino e Masculino. Se selecionarmos Feminino, o resultado será a Figura 70. Ao escolhermos Masculino obtemos a Figura 71.

Sexo		Feminino					
		Opinião					
Tipo	Dados	Excelente	Confiável	Mediana	Sofrível	Terrível	Total geral
Pós-pago	Frequências	51	39	183	170	266	709
	% por linha	7,19%	5,50%	25,81%	23,98%	37,52%	100,00%
Pré-pago	Frequências	202	388	506	235	201	1532
	% por linha	13,19%	25,33%	33,03%	15,34%	13,12%	100,00%
Total Frequências		253	427	689	405	467	2241
Total % por linha		11,29%	19,05%	30,75%	18,07%	20,84%	100,00%

Figura 70 - Tipo x Opinião para o sexo Feminino

Sexo		Masculino					
		Opinião					
Tipo	Dados	Excelente	Confiável	Mediana	Sofrível	Terrível	Total geral
Pós-pago	Frequências	86	63	311	348	576	1384
	% por linha	6,21%	4,55%	22,47%	25,14%	41,62%	100,00%
Pré-pago	Frequências	166	274	456	247	208	1351
	% por linha	12,29%	20,28%	33,75%	18,28%	15,40%	100,00%
Total Frequências		252	337	767	595	784	2735
Total % por linha		9,21%	12,32%	28,04%	21,76%	28,67%	100,00%

Figura 71 - Tipo x Opinião para o sexo Masculino

Repare que na Figura 70 o total geral vale 2241 (e não mais 4983 da Figura 68), pois são consideradas apenas as clientes. Na Figura 71, apenas os 2735 homens. A soma resulta 4976 (ao invés de 4983) porque as células vazias não foram selecionadas em Tipo, Opinião e Sexo. Podemos chegar a algumas conclusões:

- lembre-se que os clientes do Pós-pago estão mais insatisfeitos;

- mas os percentuais de insatisfação são maiores entre os homens, do que em relação ao geral – na Figura 61 pode-se ver que Sofrível e Terrível são 20,09% e 25,11% do total, e os respectivos percentuais dos homens são maiores (21,76% e 28,67%) e os das mulheres menores (18,07% e 20,84%);

- quando segmentamos em função do Tipo percebe-se o mesmo padrão anterior (Pós-pago mais insatisfeito), mas com menor variação entre homens e mulheres;

No total, em Pós-pago, há 24,74% de Sofrível e 40,13% de Terrível;

No Feminino, Pós-pago, há 23,98% de Sofrível e 37,52% de Terrível (ligeiramente abaixo do total de Pós-pago), já no Masculino há 25,14% de Sofrível e 41,62% de Terrível (ligeiramente acima do total de Pós-pago).

No total, em Pré-pago, há 16,71% de Sofrível e 14,78% de Terrível;

No Feminino, Pré-pago, há 15,34% de Sofrível e 13,12% de Terrível (ligeiramente abaixo do total de Pré-pago), já no Masculino há 18,28% de Sofrível e 15,40% de Terrível (ligeiramente acima do total de Pós-pago).

Conclui-se, então que homens são em geral mais insatisfeitos com a operadora do que as mulheres, e homens que são clientes de pós-pago são os mais descontentes. Mas, a diferença entre homens e mulheres não é tão grande: não chega a 5% em nenhuma das comparações (veja as Figura 66 a 70) indicando que a relação entre Tipo e Opinião parece ser mais forte, com pouca influência do Sexo do cliente.

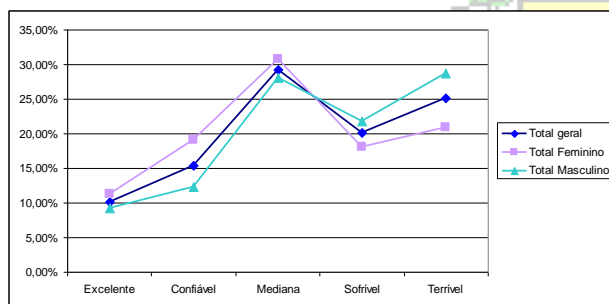


Figura 72 - Percentuais de Opinião x Sexo

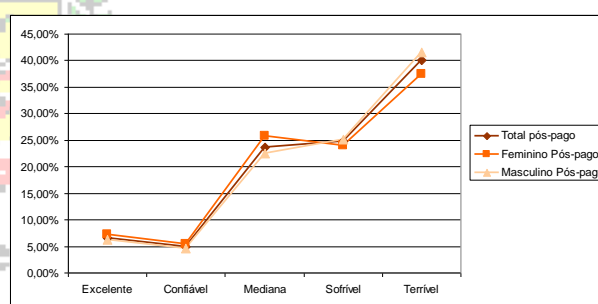


Figura 73 - Percentuais de Opinião x Sexo – Pós-pago

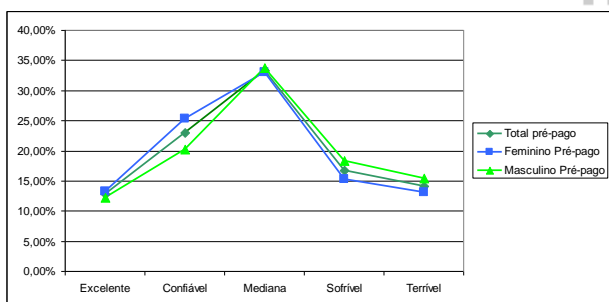


Figura 74 - Percentuais de Opinião x Sexo – Pré-pago

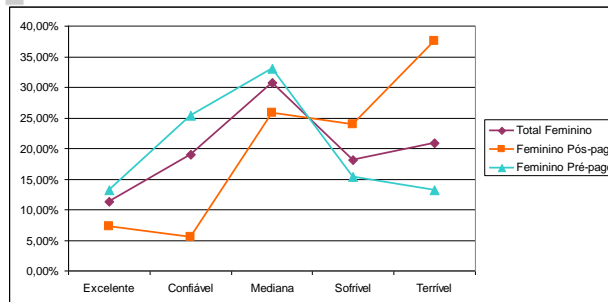


Figura 75 - Percentuais de Opinião x Tipo – Feminino

Quando segmentamos a opinião apenas por sexo (analisando o geral, apenas os clientes do pós-pago ou do pré-pago), as linhas (que representam os % para cada opinião) estão bem próximas (quase coincidentes em alguns casos), indicando comportamento similar: parece que o Sexo não é uma variável que influencie muito a opinião dos clientes, embora os homens estejam ligeiramente mais insatisfeitos – a linha dos percentuais dos homens está abaixo da do total para opiniões positivas e acima para as opiniões negativas (ver Figuras 72, 73 e 74). Mas, ao segmentar a opinião por tipo (analisando os resultados para o sexo feminino e masculino), observa-se um comportamento semelhante ao obtido na Figura 67, com a linha do pós-pago substancialmente abaixo do total para mulheres e homens para as opiniões positivas e acima para as opiniões negativas (17% acima na opinião Terrível Feminino, e 13% acima na opinião Terrível Masculino), embora eles estejam ligeiramente mais insatisfeitos do que elas (ver Figuras 75 e 76).

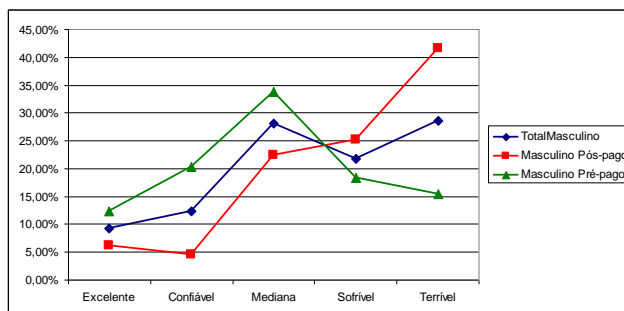


Figura 76 - Percentuais de Opinião x Tipo – Masculino

Conclui-se então que o Tipo de serviço parece ter relação mais forte com a Opinião do que o Sexo dos clientes.

4.4 – Gráficos para variáveis qualitativas

É bastante comum construir gráficos para representar distribuições de freqüências de variáveis qualitativas (uma ou mais). Os gráficos mais apropriados são os gráficos de barras (ou de colunas) e os gráficos em setores (circular ou em pizza).

4.4.1 – Gráficos para uma variável qualitativa

Suponha que há interesse em construir um gráfico para a tabela da variável qualitativa Tipo, vista na seção 4.1, Figura 48. As opções possíveis são gráfico de barras (colunas no Excel) e em setores.

Gráfico de barras (colunas)

Para criar o gráfico basta selecionar a tabela e pressionar o botão direito do mouse, como mostrado na Figura 77. A ação padrão do Excel é construir um gráfico de colunas, tal como mostrado na Figura 78.

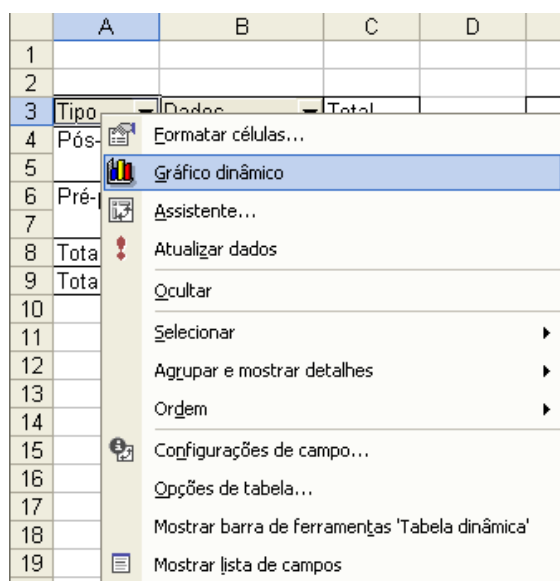


Figura 77 – Opções de tabela dinâmica: Gráfico dinâmico

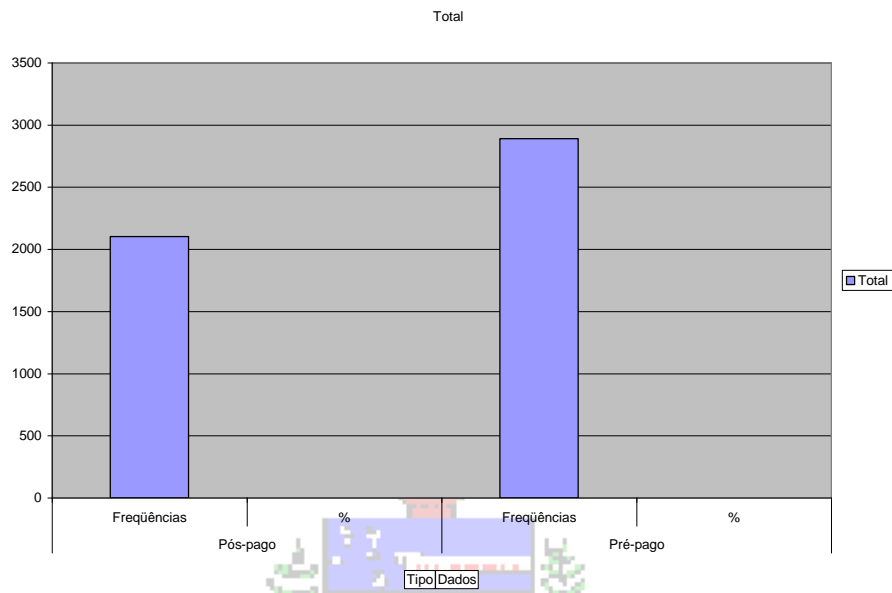


Figura 78 - Gráfico dinâmico de Tipo - inicial

Observe que o gráfico dinâmico apresenta as frequências E os percentuais (que não aparecem por serem muito menores do que as frequências. Não faz sentido manter ambos (poderíamos optar por manter apenas os percentuais), vamos desmarcar a opção referente aos percentuais, basta selecionar “Dados” na Figura 78, resultando na Figura 79. Desmarcando %, vamos obter a Figura 80.

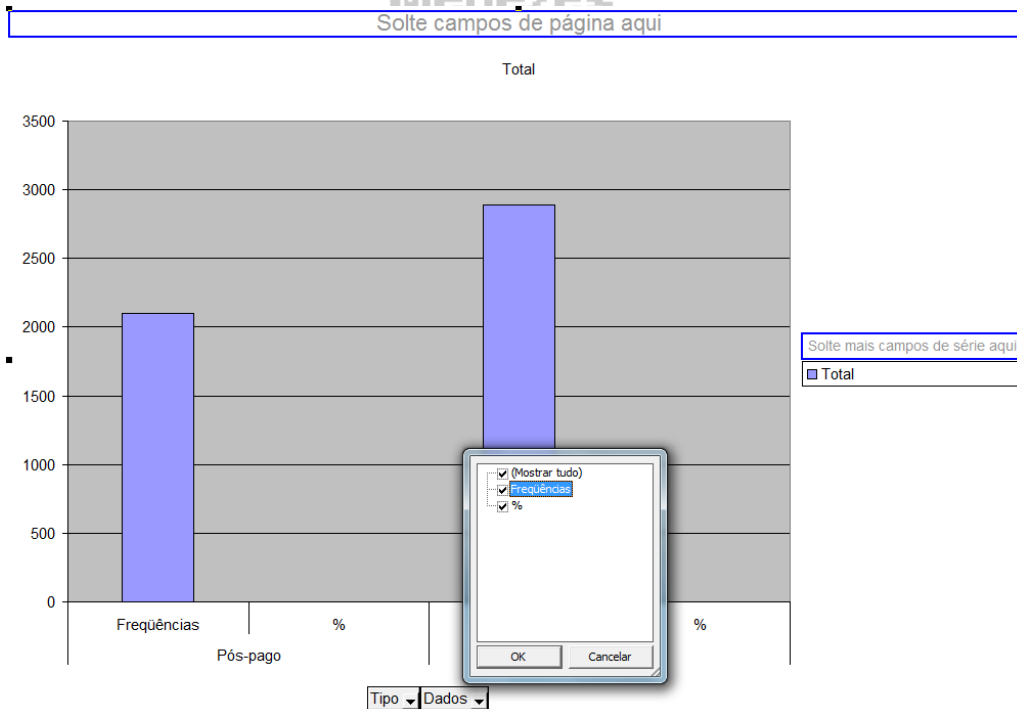


Figura 79 - Gráfico dinâmico de Tipo - Opções de apresentação de Dados

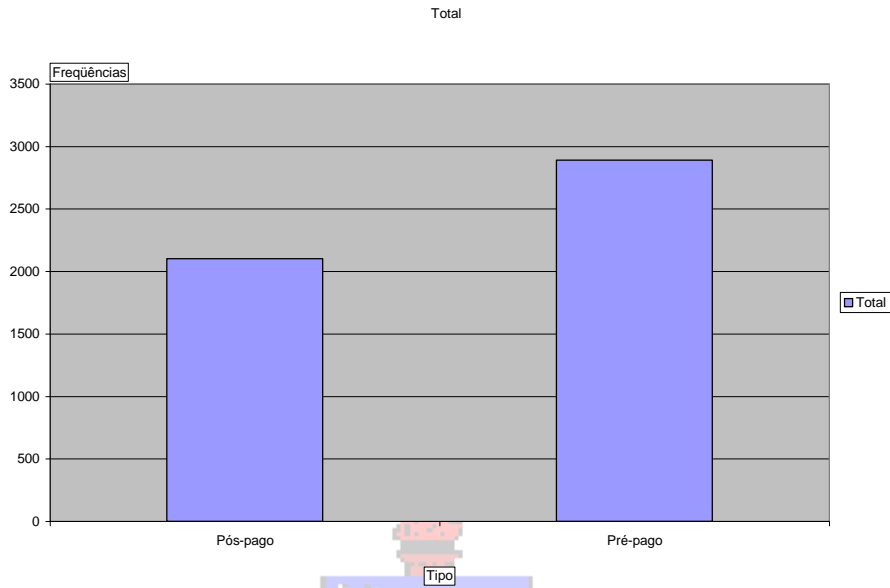


Figura 80 - Gráfico dinâmico de Tipo de serviço de telefonia celular - apenas frequências

Observe que aparecem os botões de Frequências e Tipo, além da legenda. Tudo isso precisa ser removido. Para tanto, devemos selecionar o gráfico, pressionar o botão direito do mouse, e escolher “Ocultar botões de campos de gráfico dinâmico”, na Figura 81, e o resultado surge na Figura 82.

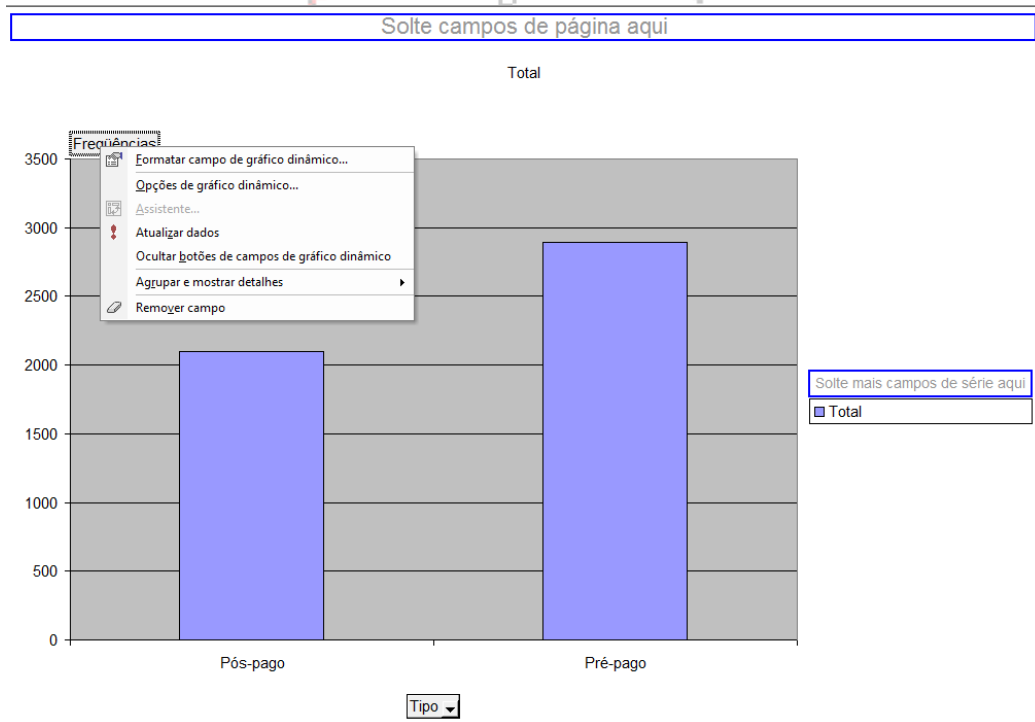


Figura 81 - Gráfico dinâmico de Tipo de serviço de telefonia celular - opções

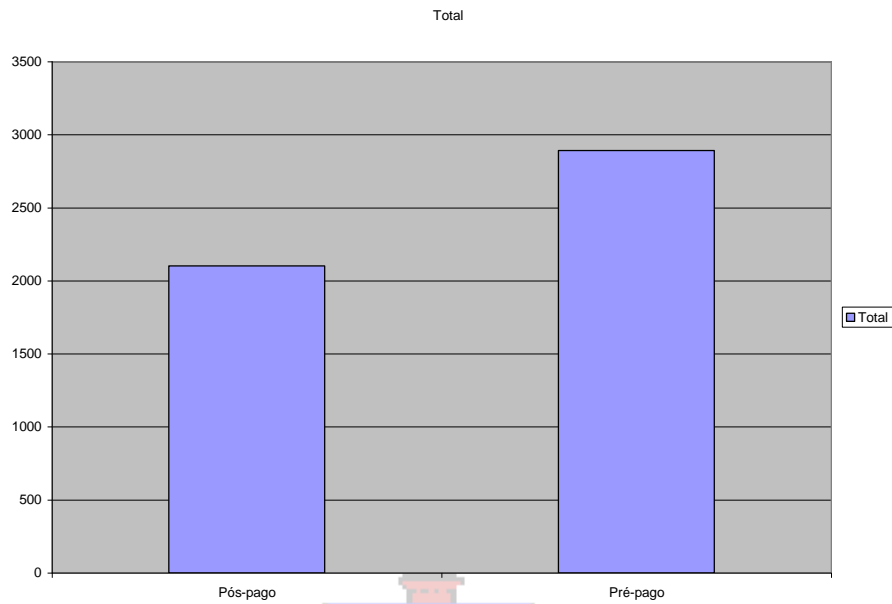


Figura 82 - Gráfico dinâmico em colunas de Tipo de serviço de telefonia celular - sem campos de gráfico dinâmico

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 Tipo. Para modificar título geral basta clicar sobre ele e reescrevê-lo: Tipo de serviço. Se selecionarmos o gráfico e clicarmos o botão direito do mouse surge a tela da Figura 83.

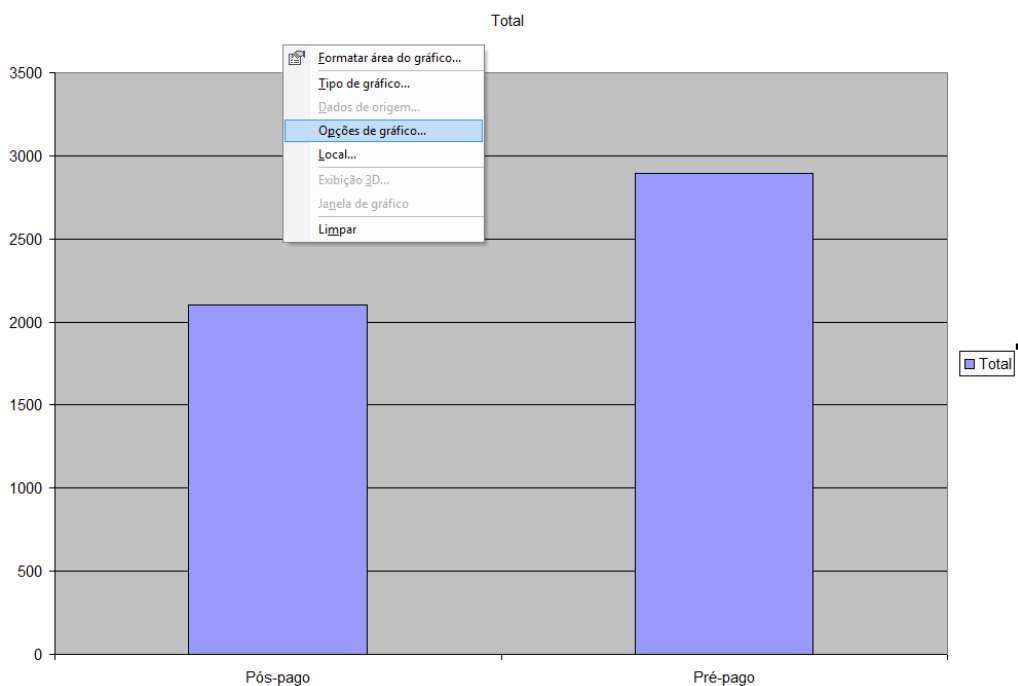


Figura 83 - Gráfico dinâmico de Tipo de serviço de telefonia celular - opções de gráfico dinâmico

Podemos mudar o tipo de gráfico (para em setores, por exemplo), formatar a área do gráfico (para tornar o fundo branco, por exemplo), mudar o Local (o Excel 2003 coloca o gráfico dinâmico diretamente em uma nova planilha, isso pode ser corrigido se quisermos), ou selecionar “Opções de gráfico”, resultando na Figura 84.

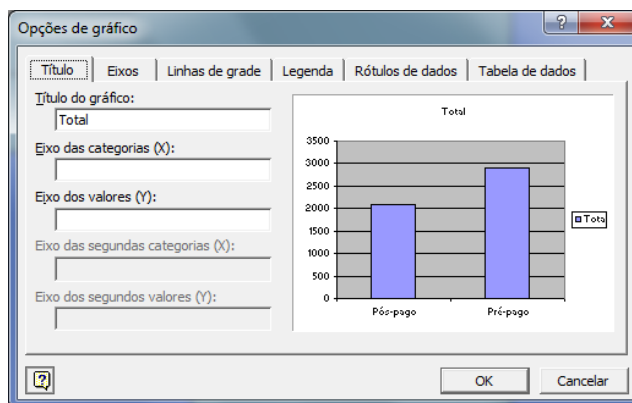


Figura 84 - Opções de gráfico

Podemos reescrever o Título do gráfico (mudar para Tipo de serviço), e o dos eixos: vamos fazer isso apenas para o eixo Y (Frequências). Se escolhermos a aba “Legenda”, a opção “Mostrar legenda” pode ser desmarcada, o que resulta no gráfico da Figura 85.

Tipo de serviço

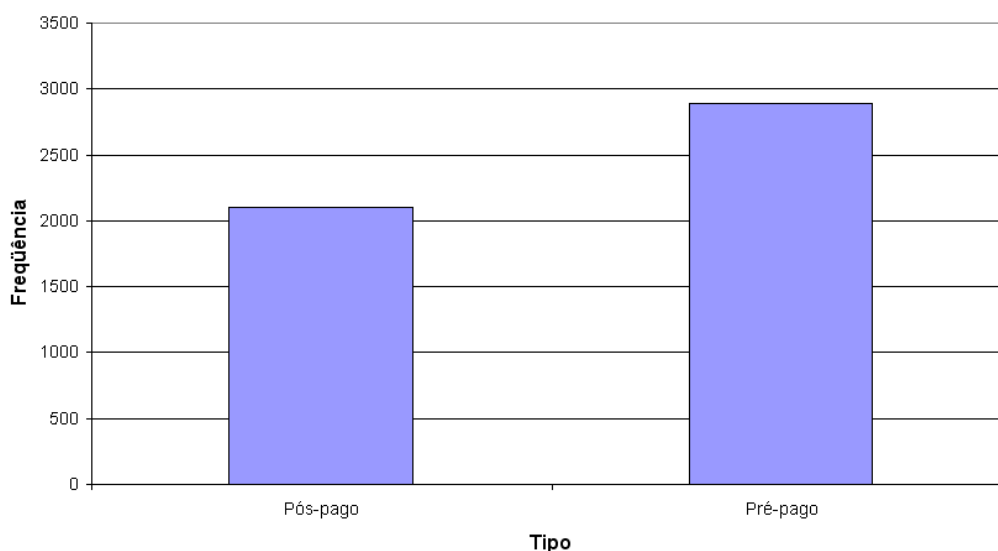


Figura 85 - Gráfico de colunas de Tipo de serviço - Final

A interpretação do gráfico é a mesma da tabela, feita na seção 4.1. 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 81 percebe-se uma diferença entre os dois tipos de serviço, com preponderância do Pré-pago (57%) sobre o Pós-pago (43%): 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 86.

Pressionando “Formatar Eixo” nos leva à Figura 87. Automaticamente o Excel 2003 escolhe a escala mais apropriada de acordo com os dados de origem (Escala do eixo dos valores: Automática). Mas, podemos trocar os limites e as unidades da escala. Na Figura 88 trocamos o mínimo para Fixo, indicando que o eixo vertical começará em 1000. O gráfico resultante é mostrado na Figura 89.

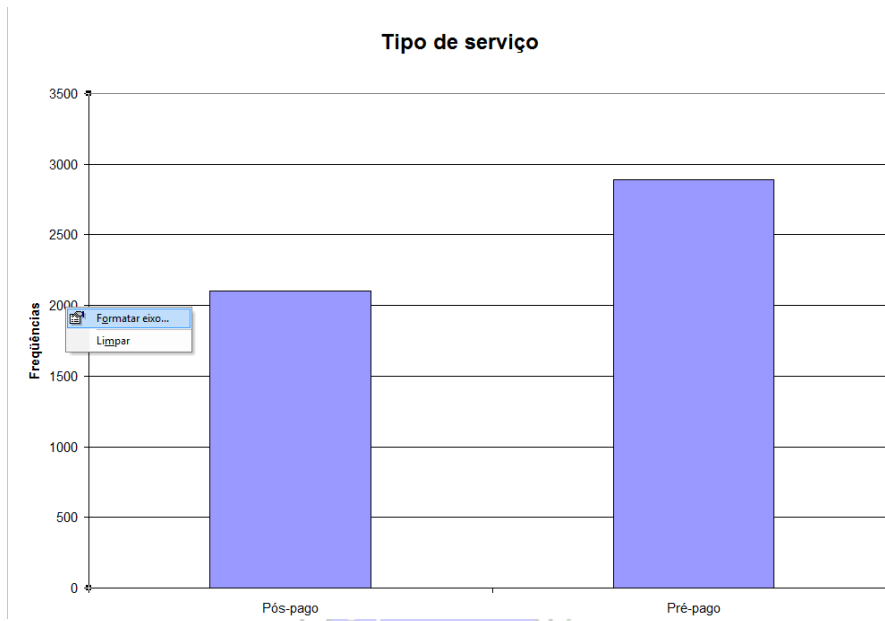


Figura 86 - Gráfico em colunas – opções de eixo

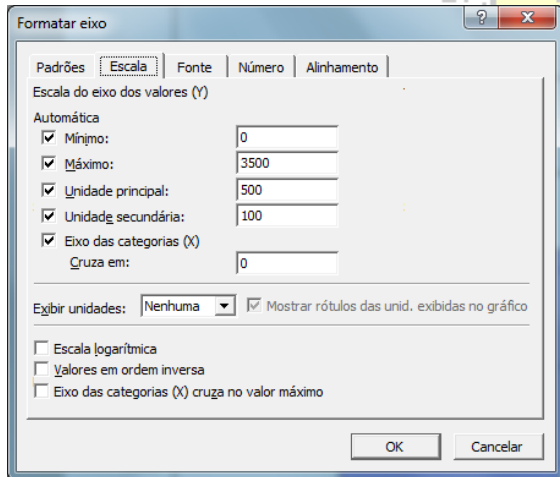


Figura 87 - Gráfico em colunas – escala

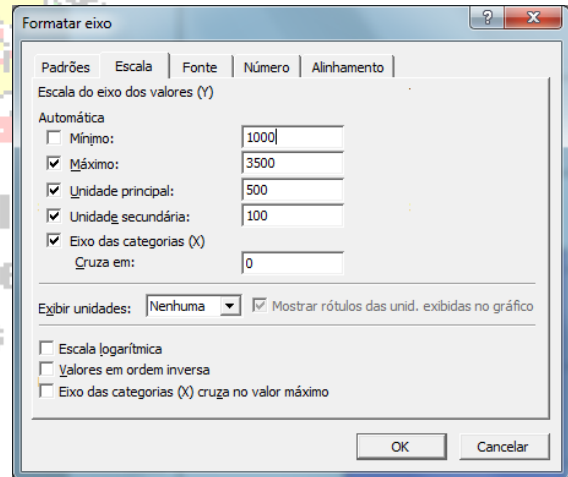


Figura 88 - Mudança de escala no eixo vertical

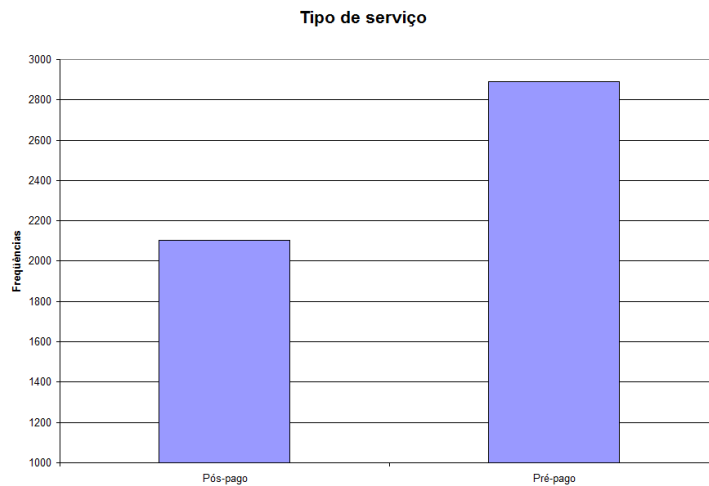


Figura 89 - Gráfico de colunas de Tipo de serviço – escala vertical modificada

Observe que a diferença entre os dois tipos de serviço 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. Para criar o gráfico basta selecionar a tabela e pressionar o botão direito do mouse, como mostrado na Figura 77. A ação padrão do Excel é construir um gráfico de colunas, tal como mostrado na Figura 78. Vamos modificá-lo através das opções mostradas na Figura 83, escolhendo Tipo de Gráfico, o que resultará na tela à esquerda da Figura 90.

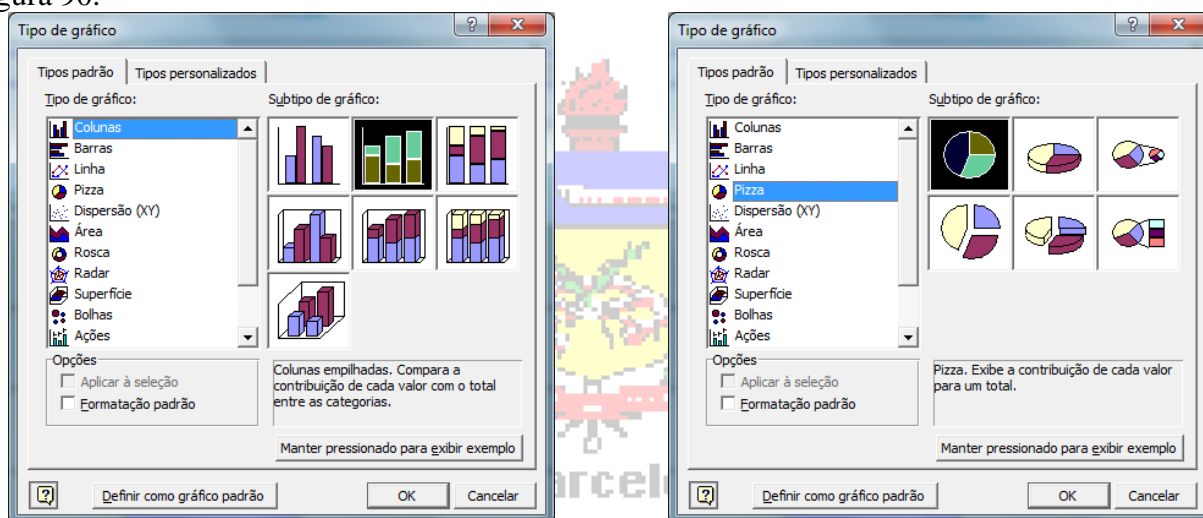


Figura 90 - Modificação no tipo de gráfico dinâmico

Escolhendo Pizza, 2D, o resultado é o gráfico da Figura 91 (já com os botões de gráfico dinâmico removidos – ver Figura 81).

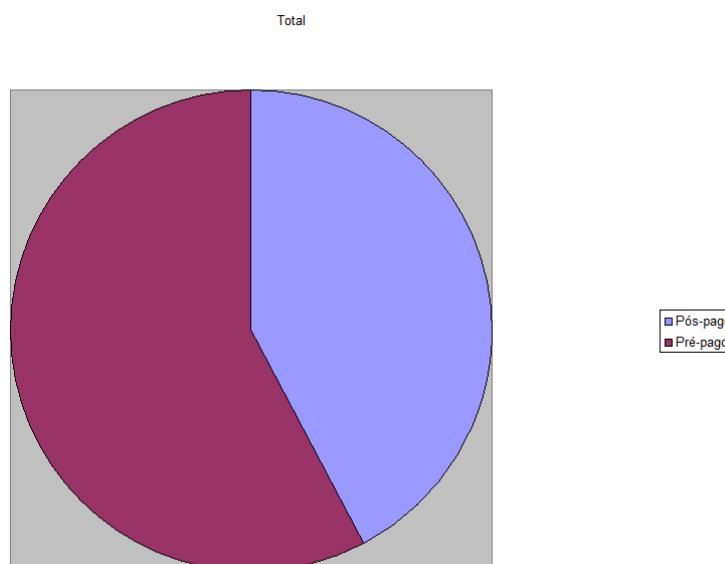


Figura 91 - Gráfico dinâmico em pizza de Tipo de serviço de telefonia celular - sem campos de gráfico dinâmico

Observe que não há percentuais, o que prejudica a interpretação do gráfico. Selecionando o gráfico e pressionando o botão direito do mouse temos a mesma situação da Figura 83. Escolhendo “Opções de gráfico” as possibilidades específicas para o gráfico em Pizza surgem na Figura 92.

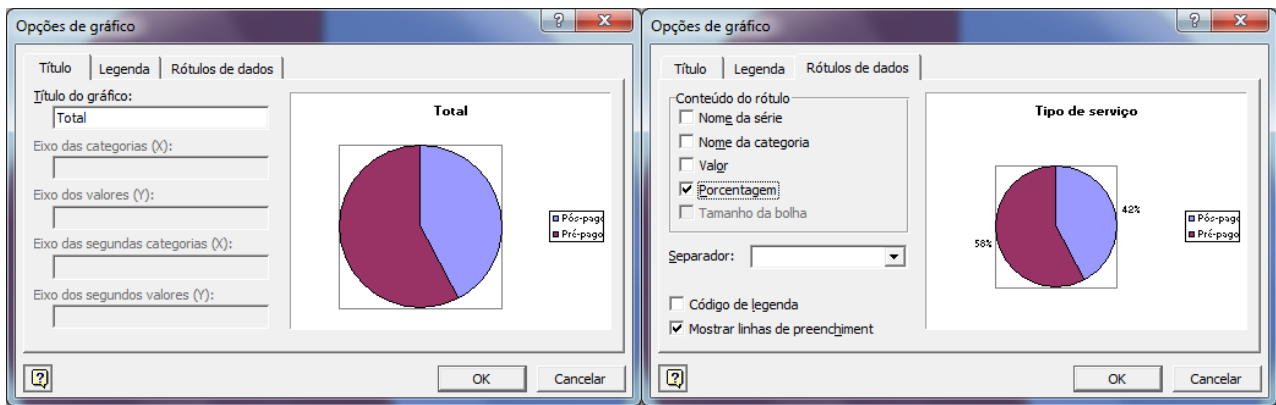


Figura 92 - Opções de Gráfico - gráfico em pizza

Na aba “Título” podemos mudar o título geral do gráfico. Na aba “Legenda” a sua posição pode ser deslocada, e em “Rótulos de dados” é possível marcar “Porcentagem”, para que o gráfico fique completo (Figura 93).



Tipo de serviço

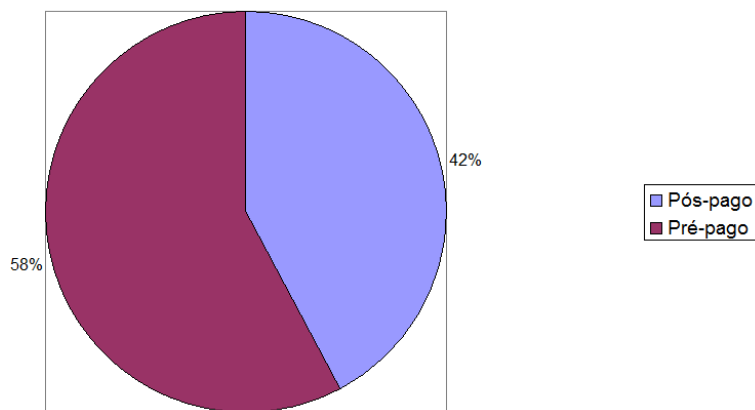


Figura 93 - Gráfico em pizza de Tipo de Serviço - Final

4.4.2 – Gráficos para duas variáveis qualitativas

Suponha que há interesse em construir um gráfico para a tabela conjunta das variáveis qualitativas Tipo e Opinião, vista na seção 4.2, Figura 66. A opção é gráfico de barras (colunas no Excel), mas podendo usar vários subtipos. Antes, porém, vamos copiar a tabela dinâmica inteira para uma área livre na planilha e remover todos os percentuais, o que resultará na Figura 94.

Opinião	Excelente	Confiável	Mediana	Sofrível	Terrível	Total geral
Tipo						
Pós-pago	137	103	497	519	842	2098
Pré-pago	368	664	962	482	409	2885
Total geral	505	767	1459	1001	1251	4983

Figura 94 - Tabela dinâmica de Tipo por Opinião - apenas freqüências absolutas

Para criar o gráfico basta selecionar a tabela e pressionar o botão direito do mouse, como mostrado na Figura 77 para uma variável. A ação padrão do Excel é construir um gráfico de colunas empilhadas, tal como mostrado na Figura 95.

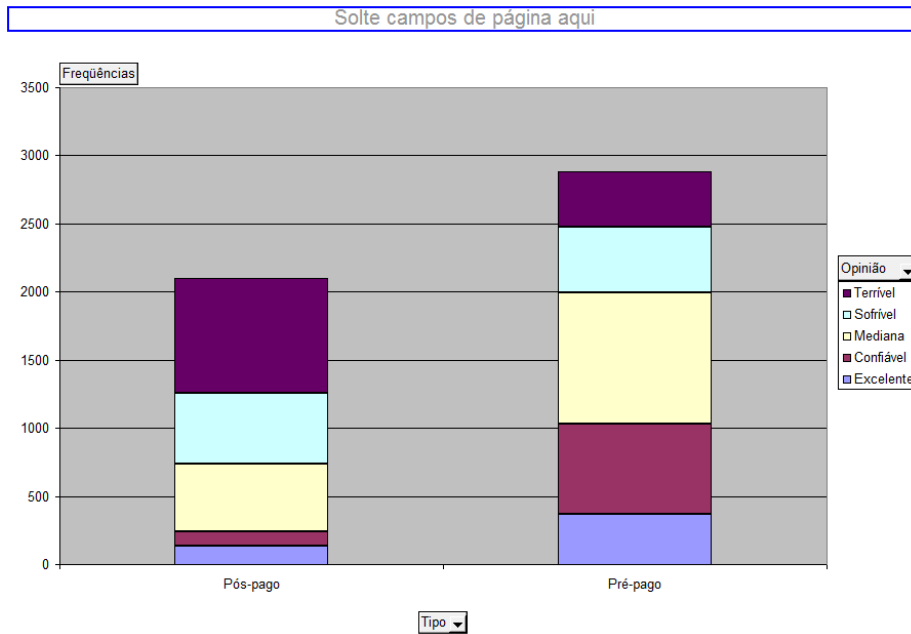


Figura 95 - Gráfico de colunas 2D empilhadas de Tipo por Opinião

Observe que não há problemas com os percentuais, nem com a escala que começa em zero. Mas, permanecem os campos de gráfico dinâmico (que já sabemos como remover), e, principalmente, o tipo de gráfico torna difícil a interpretação. Se selecionarmos o gráfico da Figura 95 e pressionarmos o botão direito do mouse o resultado será algo parecido com a Figura 83: e selecionando “Tipo de Gráfico” chegamos à tela esquerda na Figura 96.

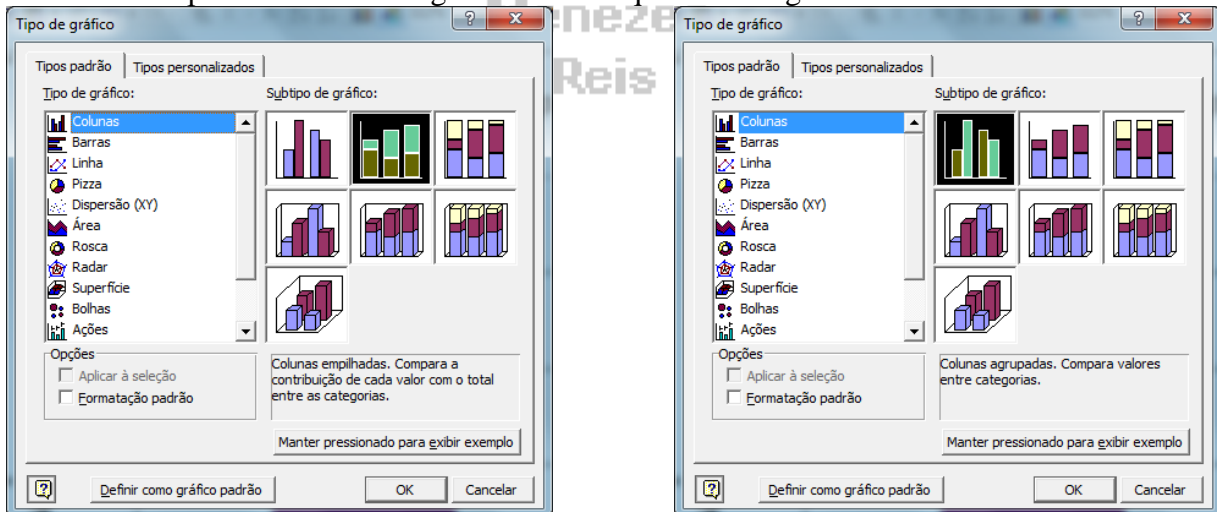


Figura 96 - Mudança de tipo de gráfico de colunas: empilhadas para agrupadas

Podemos mudar para colunas agrupadas (como mostrado à direita na Figura 96), o que resultaria na Figura 97 (já com mudanças nos títulos, plano de fundo e cores das barras, e com os campos dinâmicos removidos). Ou, como vamos ver mais à frente, poderíamos escolher colunas 100% empilhadas.

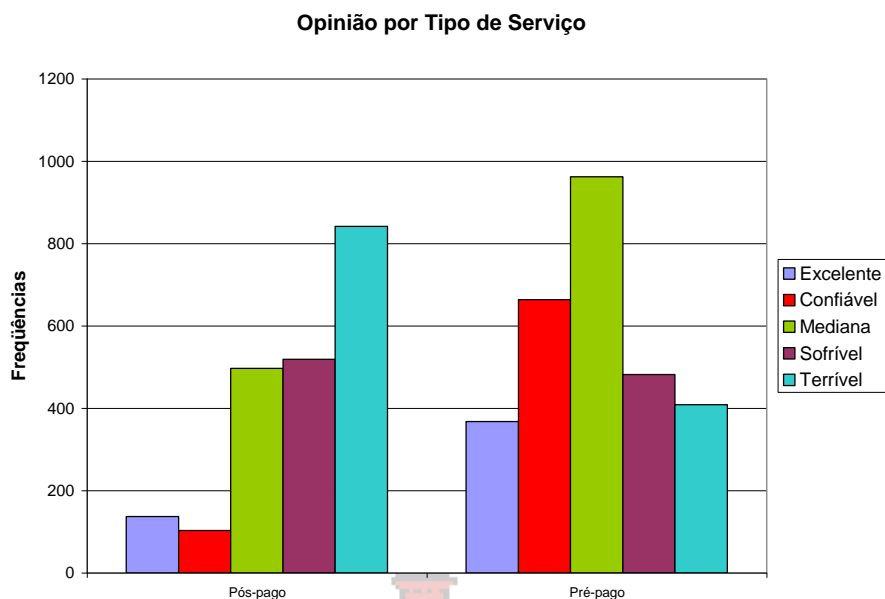


Figura 97 - Gráfico de colunas 2D agrupadas de Tipo por Opinião

Percebe-se que as frequências de “Excelente” e “Confiável” são maiores em Pré-pago do que em Pós-pago, e, as de “Sofrível” e “Terrível” são maiores neste último. Para interpretar este gráfico é preciso olhar para a “altura” e disposição das barras: quando elas são diferentes, como no caso acima, indicam relacionamento entre as variáveis (dependendo do Tipo de serviço a satisfação é maior). Mas, quando elas são semelhantes, isso significa que as variáveis apresentam pouco ou nenhum relacionamento. Imagine que a tabela dinâmica de Tipo por Opinião levasse ao gráfico da Figura 98:

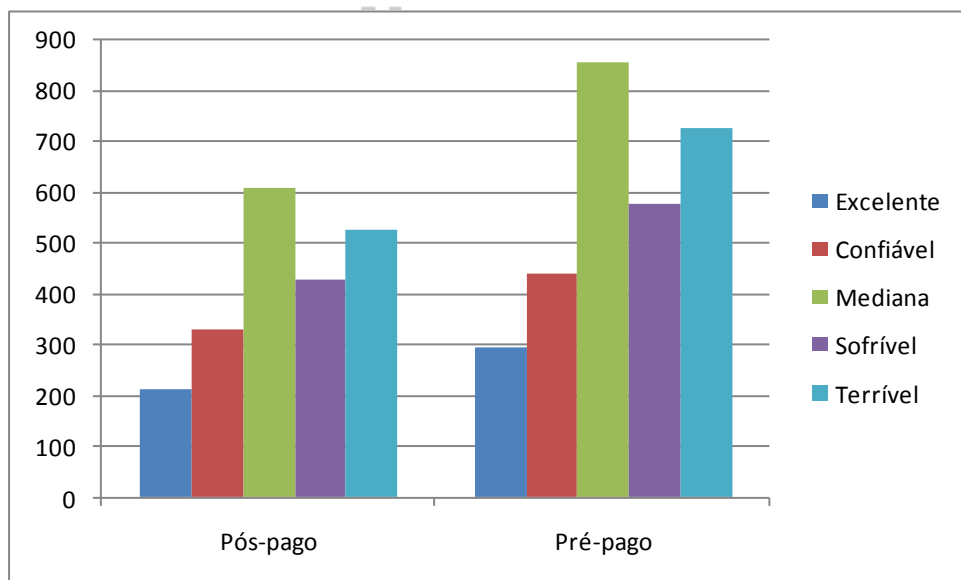


Figura 98 - Gráfico de colunas 2D agrupadas de Tipo por Opinião – Pouco relacionamento

Os dois grupos de barras são muito semelhantes, embora a quantidade de clientes do Pré-pago seja maior, é possível perceber que o comportamento é bastante semelhante.

O gráfico de colunas 2D agrupadas, no entanto, exige uma certa prática para sua correta interpretação, o que pode causar alguns problemas nos casos em que as diferenças entre as variáveis não forem grandes. Uma solução é usar outro subtipo de gráfico de colunas 2D: colunas 2D 100% empilhadas, onde serão apresentados os percentuais em relação às linhas ou colunas, o que

permitirá uma análise mais acurada. Para o gráfico da Figura 97, basta selecionar o gráfico e pressionar o botão direito do mouse, e procurar as opções de Tipo de gráfico, na tela à direita na Figura 96, e escolhendo “colunas 100% empilhadas” (Figura 99), chegaremos ao gráfico final, na Figura 100.

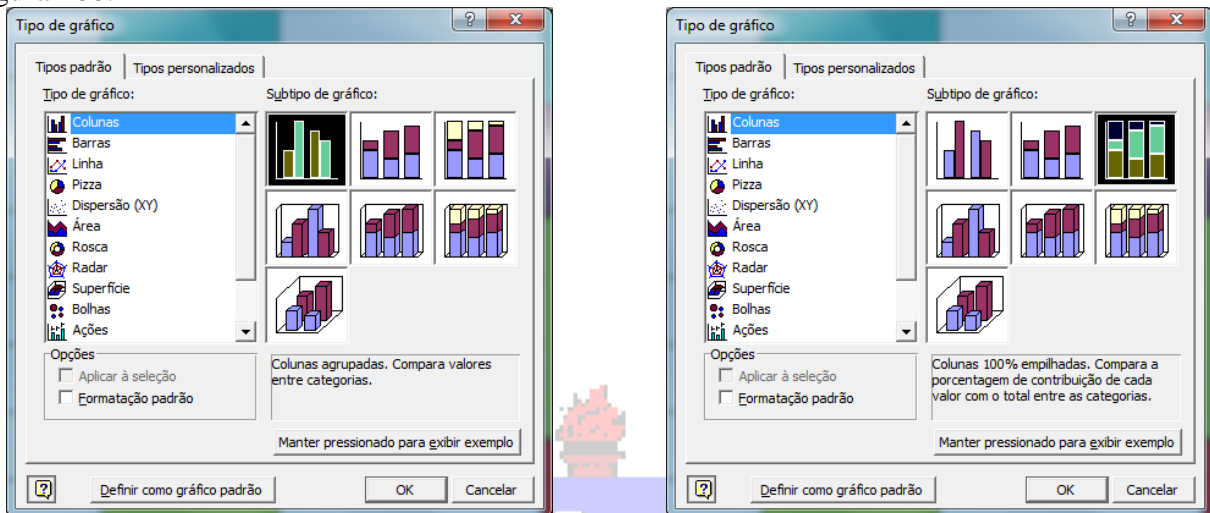


Figura 99 - Mudança de tipo de gráfico de colunas: agrupadas para 100% empilhadas

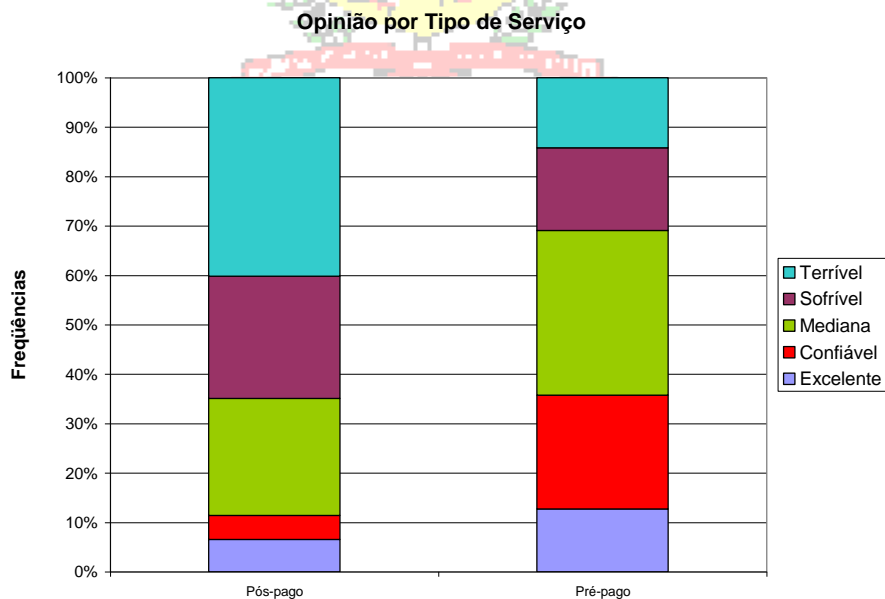


Figura 100 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião

É possível perceber claramente que os clientes do Pós-pago apresentam percentuais de Terrível e Sofrível bem maiores do que os do Pré-pago, como havíamos concluído anteriormente ao analisar os percentuais por linha na Figura 66.

Ao modificar o gráfico da Figura 98 chegamos à Figura 101:

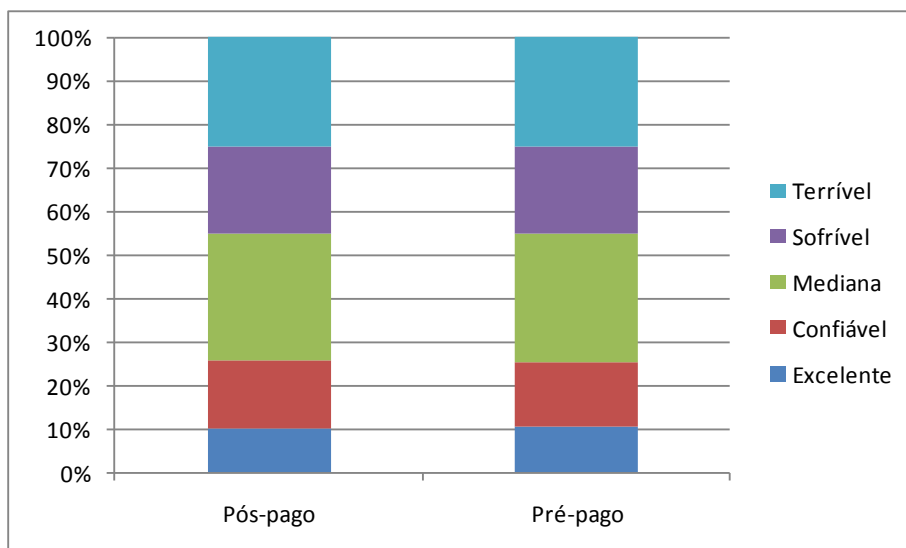


Figura 101 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião – Pouco relacionamento

Ao contrário do caso da Figura 100 as duas colunas são muito parecidas, com os percentuais para bastante próximos, o que indica pouco relacionamento entre as variáveis Tipo e Opinião.

Para fazer um gráfico com três variáveis qualitativas, basta seguir os passos desta seção usando a tabela com 3 variáveis qualitativas mostrada na seção 4.3. É possível construir um gráfico para cada valor da variável Sexo: um para Feminino (Figura 70) e outro para Masculino (Figura 71), e compará-los⁸.

Marcelo
Menezes
Reis

5. Análise de dados de 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.

5.1 - Procedimentos para variáveis quantitativas discretas – tabelas dinâmicas

Se a variável for discreta, por exemplo, a variável Anos, no arquivo Celular01.xls, que representa o número de anos completos em que cada pessoa daquele arquivo é cliente da operadora de celular. Vamos construir uma tabela dinâmica para esta variável, repetindo os procedimentos vistos para uma variável qualitativa, Figuras 33 a 39, pondo os resultados em uma planilha existente ou não. Chegamos então ao layout da tabela dinâmica de Anos, mostrado na Figura 102: a variável Anos vai para Rótulos de Linha e Valores (não incluiremos nenhum percentual ainda).

⁸ O ideal é copiar a tabela dinâmica com três variáveis, e construir os gráficos separadamente, para poder compará-los.

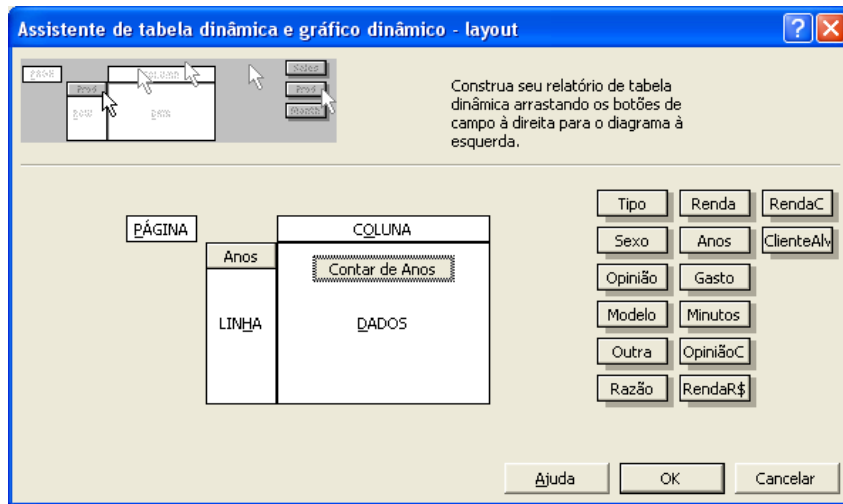


Figura 102 - Layout da tabela dinâmica de Anos

A variável Anos foi arrastada para o campo “LINHA” e para o campo “DADOS”. A operação padrão do Excel é apresentar Soma de Anos, já que esta é quantitativa. Na Figura 102, já mudamos a opção de resumo dos dados para contagem, pois queremos uma tabela de frequências. E na Figura 103 acrescentamos mais uma contagem de anos ao campo “DADOS” para calcular os respectivos percentuais.

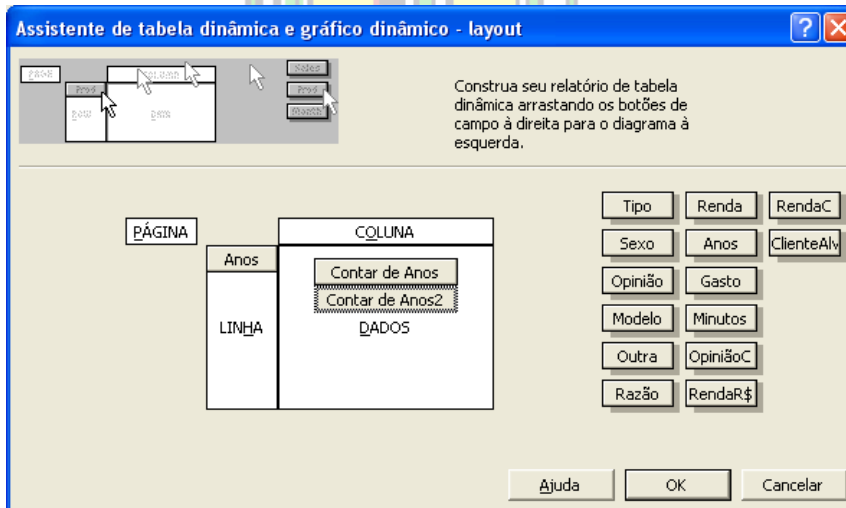


Figura 103 - Layout da Tabela dinâmica de Anos

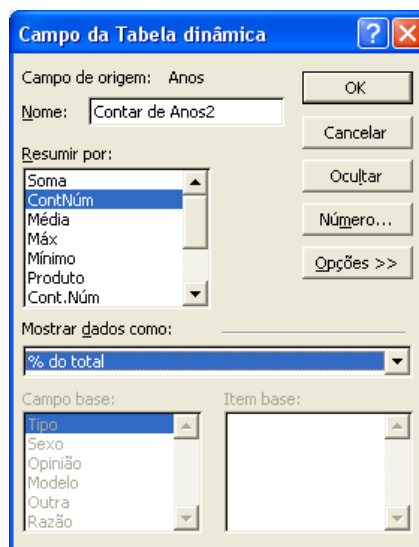


Figura 104 - Mudando a forma de mostrar os dados

Na Figura 104 repetimos o procedimento mostrado na Figura 43: a segunda contagem de Anos é modificada para ser mostrada como percentual do total da variável. Agora podemos finalizar o Layout e obter a tabela dinâmica de Anos (já sem células vazias), vista na Figura 105.

Anos	Dados	Total
0	Frequência	38
	%	0,76%
1	Frequência	154
	%	3,08%
2	Frequência	312
	%	6,24%
3	Frequência	582
	%	11,65%
4	Frequência	863
	%	17,27%
5	Frequência	1011
	%	20,24%
6	Frequência	877
	%	17,55%
7	Frequência	600
	%	12,01%
8	Frequência	358
	%	7,17%
9	Frequência	141
	%	2,82%
10	Frequência	44
	%	0,88%
11	Frequência	12
	%	0,24%
12	Frequência	2
	%	0,04%
14	Frequência	2
	%	0,04%
Total Frequência		4996
Total %		100,00%

Observe que não há o valor 13, pois não há nenhuma pessoa que seja cliente há 13 anos.

A tabela da Figura 105 não está correta do ponto de vista estatístico. Quando lidamos com uma variável quantitativa discreta, todos os valores entre os extremos precisam fazer parte da tabela, mesmo que não ele não ocorra no conjunto de dados (tenha frequência zero). Precisamos então fazer mais modificações para obter a tabela correta.

Vamos aproveitar que o Excel já computou todas as frequências dos outros valores, e copiá-los para uma nova tabela. Mas, antes de tudo, é preciso remover do campo “DADOS” os percentuais, vamos recalculá-los posteriormente. Ao desmarcar o valor % (Figura 106), resultando na tabela da Figura 107.

Para tornar a tabela correta podemos copiá-la para células apropriadas em outra parte da planilha, incluindo as frequências. Isso é mostrado na Figura 108.

O próximo passo é corrigir novamente a tabela, para incluir o valor 13, que exigirá a inserção de células entre as células referentes aos valores 12 e 14.

Figura 105 - Tabela dinâmica de Anos

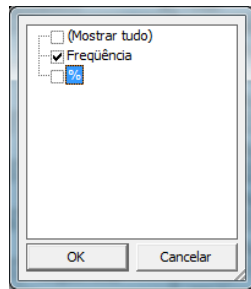


Figura 106 – Opções de apresentação de Dados

Frequência	Total
Anos	
0	38
1	154
2	312
3	582
4	863
5	1011
6	877
7	600
8	358
9	141
10	44
11	12
12	2
14	2
Total geral	4996

Figura 107 - Tabela de Anos - sem %

Frequência	Total	Anos	Frequências
Anos			
0	38	0	38
1	154	1	154
2	312	2	312
3	582	3	582
4	863	4	863
5	1011	5	1011
6	877	6	877
7	600	7	600
8	358	8	358
9	141	9	141
10	44	10	44
11	12	11	12
12	2	12	2
14	2	14	2
Total geral	4996		

Figura 108 - Cópia das células da Tabela dinâmica de Anos

	N	O	P	Q	R	S
3	Contar de Anos					
4	Anos	Total			Anos	Frequência
5		0	38		0	38
6		1	154		1	154
7		2	312		2	312
8		3	582		3	582
9		4	862		4	862
10		5	1011		5	1011
11		6	877		6	877
12		7	600		7	600
13		8	358		8	358
14		9	141		9	141
15		10	44		10	44
16		11	12		11	12
17		12	2		12	2
18		14	2		14	2
19	Total geral		4995			

Para inserir células no Excel precisamos selecionar o local apropriado: no nosso caso, entre as células referentes aos valores 12 e 14: devemos então selecionar as células correspondentes ao valor 14 anos (e sua frequência), pois o Excel sempre insere as células acima do selecionado. Posteriormente, no menu “Inserir”, opção “Células”, como mostrado na Figura 110, resultando na Figura 111.

Figura 109 - Inserção de células na Tabela de Anos - 1a parte

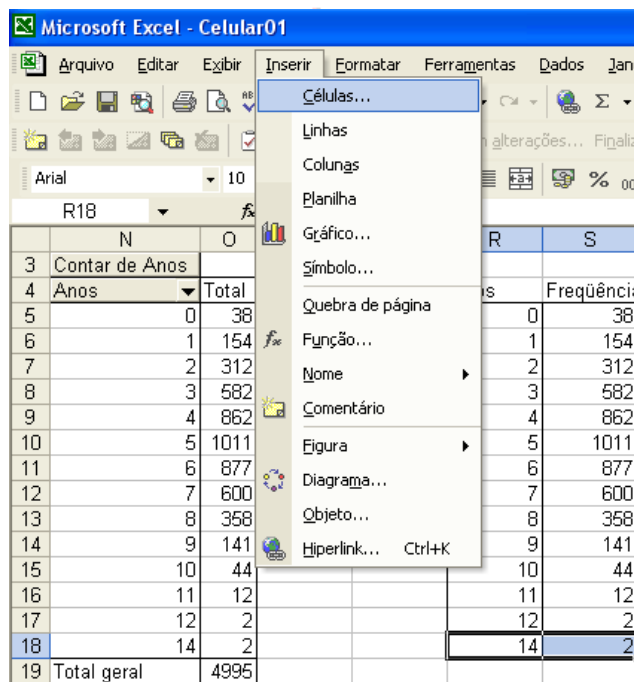


Figura 110 - Menu Inserir - Opção Células

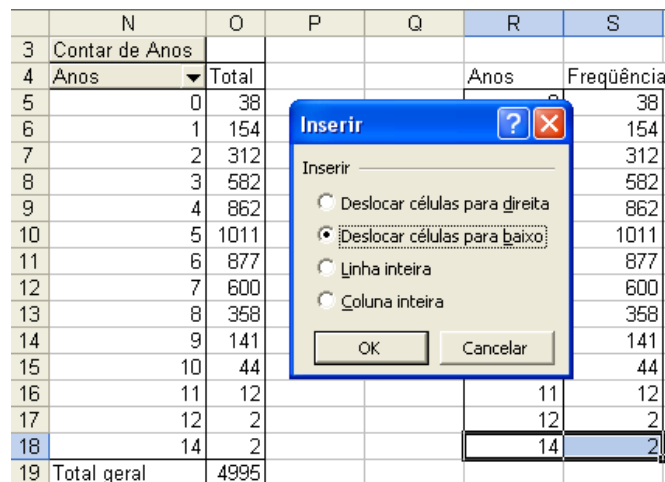


Figura 111 - Inserção de células: deslocamento para baixo

Basta escolher “Deslocar células para baixo” na Figura 111, resultando na Figura 112.

	N	O	P	Q	R	S
3	Contar de Anos					
4	Anos	Total			Anos	Frequência
5	0	38			0	38
6	1	154			1	154
7	2	312			2	312
8	3	582			3	582
9	4	862			4	862
10	5	1011			5	1011
11	6	877			6	877
12	7	600			7	600
13	8	358			8	358
14	9	141			9	141
15	10	44			10	44
16	11	12			11	12
17	12	2			12	2
18	14	2				
19	Total geral	4995			14	2

Agora precisamos completar a tabela com o valor 13 e sua frequência, que vale zero.

Ao fazermos isso completamos a tabela de frequências de Anos (Figura 113).

O último passo é calcular os percentuais, resultando na Figura 114.

Figura 112 - Células inseridas na Tabela de Anos

	N	O	P	Q	R	S
3	Contar de Anos					
4	Anos	Total			Anos	Frequência
5	0	38			0	38
6	1	154			1	154
7	2	312			2	312
8	3	582			3	582
9	4	862			4	862
10	5	1011			5	1011
11	6	877			6	877
12	7	600			7	600
13	8	358			8	358
14	9	141			9	141
15	10	44			10	44
16	11	12			11	12
17	12	2			12	2
18	14	2			13	0
19	Total geral	4995			14	2

Figura 113 - Tabela de frequências de Anos

	N	O	P	Q	R	S	T
3	Anos	Dados	Total				
4	0	Contar de Anos	38		Anos	Frequência	%
5		Contar de Anos2	0,76%		0	38	0,76%
6	1	Contar de Anos	154		1	154	3,08%
7		Contar de Anos2	3,08%		2	312	6,25%
8	2	Contar de Anos	312		3	582	11,65%
9		Contar de Anos2	6,25%		4	862	17,26%
10	3	Contar de Anos	582		5	1011	20,24%
11		Contar de Anos2	11,65%		6	877	17,56%
12	4	Contar de Anos	862		7	600	12,01%
13		Contar de Anos2	17,26%		8	358	7,17%
14	5	Contar de Anos	1011		9	141	2,82%
15		Contar de Anos2	20,24%		10	44	0,88%
16	6	Contar de Anos	877		11	12	0,24%
17		Contar de Anos2	17,56%		12	2	0,04%
18	7	Contar de Anos	600		13	0	0,00%
19		Contar de Anos2	12,01%		14	2	0,04%
20	8	Contar de Anos	358		Total	4995	100%
21		Contar de Anos2	7,17%				
22	9	Contar de Anos	141				
23		Contar de Anos2	2,82%				
24	10	Contar de Anos	44				
25		Contar de Anos2	0,88%				
26	11	Contar de Anos	12				
27		Contar de Anos2	0,24%				
28	12	Contar de Anos	2				
29		Contar de Anos2	0,04%				
30	14	Contar de Anos	2				
31		Contar de Anos2	0,04%				
32	Total Contar de Anos		4995				

Figura 114 - Tabela de Anos: frequências e percentuais

5.1.1 – Distribuição de frequência acumulada

Quando lidamos com variáveis quantitativas (discretas ou contínuas) torna-se interessante acrescentar à tabela as frequências e percentuais acumulados, o que pode enriquecer a análise. Para tanto podemos acrescentar à tabela da Figura 114 uma coluna, e usar a função CONT.SE, que é pré-programada no Excel, e pode ser usada para qualquer tipo de variável. Esta função conta quantos valores em um determinado intervalo de dados atendem a um critério estabelecido. O estabelecimento do critério não permite, porém, que sejam incluídos dois limites, por exemplo, contar todos os valores que são maiores do que 15000 e menores do que 27000. Mas, é possível utilizar outras células da planilha ao definir o critério (por exemplo, contar todos os valores menores do que o conteúdo da célula L12). Esta função tem a seguinte sintaxe:

CONT.SE(intervalo de células; critérios de contagem)

Sabe-se que os dados da variável Anos encontram-se nas células H2 a H5001 da planilha de dados⁹, este deverá ser o intervalo inserido na função CONT.SE. O critério deverá ser “igual ao conteúdo” da célula R5, que contém o primeiro valor da variável (Pós-pago). Vejam na Figura 115:

	A	B	C	D	E	F	G	H
1								
2								
3	Frequência							
4	Anos	Total		Anos	Frequências	%	Freq.acumulada	
5	0	38		0	38	0,761%	=CONT.SE(Plan1!\$H\$2:\$H\$5001;"<="&D5)	
6	1	154		1	154	3,082%		
7	2	312		2	312	6,245%		
8	3	582		3	582	11,649%		
9	4	863		4	863	17,274%		
10	5	1011		5	1011	20,236%		
11	6	877		6	877	17,554%		
12	7	600		7	600	12,010%		
13	8	358		8	358	7,166%		
14	9	141		9	141	2,822%		
15	10	44		10	44	0,881%		
16	11	12		11	12	0,240%		
17	12	2		12	2	0,040%		
18	14	2		13	0	0,000%		
19	Total geral	4996		14	2	0,040%		
20				Total	4996	100%		

Figura 115 - CONT.SE: primeira frequência acumulada da variável Ano

O critério *precisa* ser colocado entre aspas (“<=”), significando que todos os valores menores ou iguais à referência serão contabilizados naquela célula. A referência à célula D5 precisa ser precedida pelo caracter especial &, ou a função retornará uma mensagem de erro. Observe que o intervalo H2:H5001 tem alguns caracteres \$. Isso é feito para que este intervalo torne-se uma referência absoluta, para quando arrastarmos a fórmula até o fim da tabela, a contagem seja feita sempre no mesmo intervalo. O resultado final está na Figura 116.

⁹ No arquivo Celular01.xls é a Plan1.

	D	E	F	G
4	Anos	Freqüências	%	Freq.acumulada
5	0	38	0,761%	38
6	1	154	3,082%	192
7	2	312	6,245%	504
8	3	582	11,649%	1086
9	4	863	17,274%	1949
10	5	1011	20,236%	2960
11	6	877	17,554%	3837
12	7	600	12,010%	4437
13	8	358	7,166%	4795
14	9	141	2,822%	4936
15	10	44	0,881%	4980
16	11	12	0,240%	4992
17	12	2	0,040%	4994
18	13	0	0,000%	4994
19	14	2	0,040%	4996
20	Total	4996	100%	

Figura 116 - Tabela de Anos: freqüências absolutas, acumuladas e %

Observe-se que 95,98% das pessoas são clientes a no máximo 8 anos da operadora. E que apenas 21,74% tornaram-se clientes nos últimos 3 anos, corroborando a análise feita acima.

A avaliação cuidadosa das freqüências e percentuais acumulados permite obter muitas informações preciosas sobre os dados. A freqüência e o percentual acumulados também podem ser apresentados na direção oposta, dos maiores para os menores valores, se for do interesse do analista.

5.1.2 – Histograma para variáveis quantitativas discretas

O gráfico de colunas para uma variável *quantitativa* recebe o nome especial de histograma. Trata-se de um gráfico de colunas justapostas (com espaçamento nulo entre elas) em que as áreas das colunas são proporcionais às freqüências dos valores. Vamos usar a tabela que construímos para a variável Anos (Figura 116). Acionando o assistente gráfico surge a Figura 117.

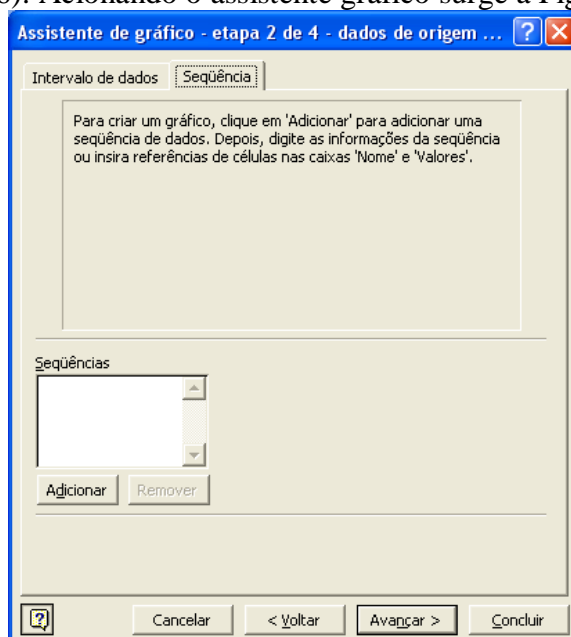


Figura 117 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa

Precisamos adicionar os dados então pressionamos “Adicionar” na Figura 117, resultando na Figura 118. No campo “Valores” devemos pôr as freqüências (Figura 119).

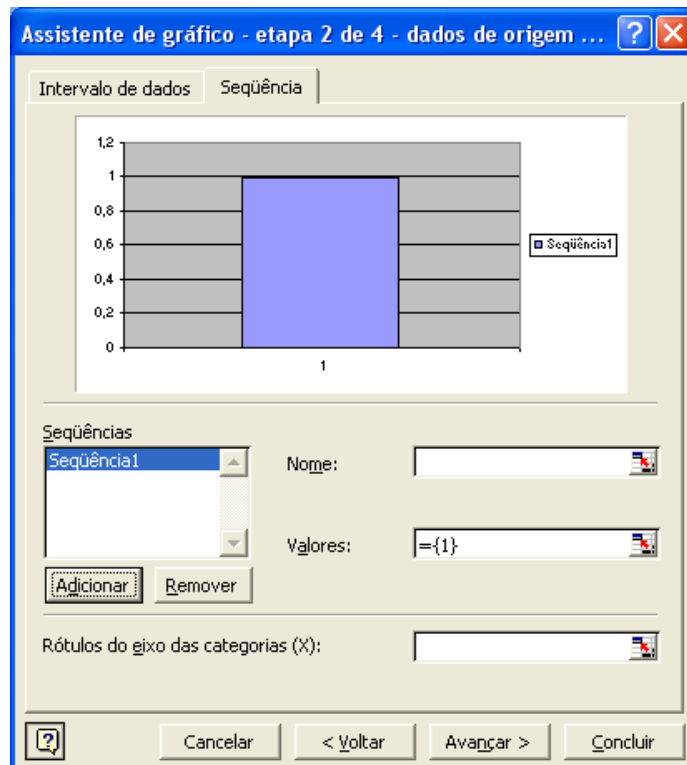


Figura 118 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – seleção de valores

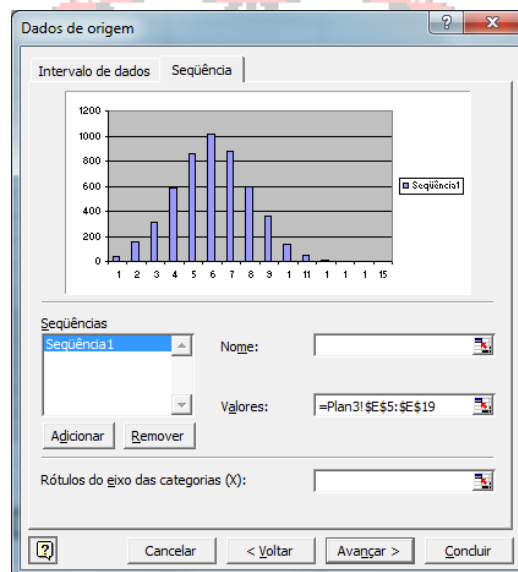


Figura 119 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – valores adicionados

Observe na Figura 119 que os números no eixo X não correspondem corretamente aos valores da variável Anos. Eles precisam ser adicionados como “Rótulos do eixo das categorias (X), como na Figura 120.

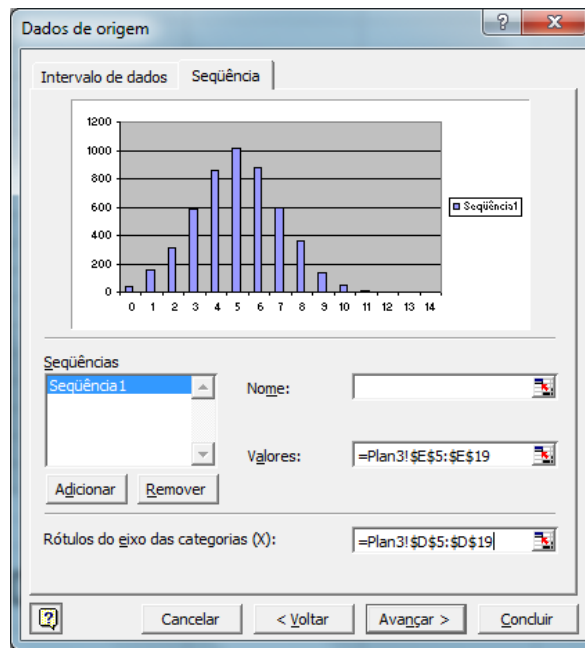


Figura 120 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – seleção de rótulos do eixo X

Observe na Figura 120 que a correspondência é correta. Podemos agora acrescentar título, remover legenda, etc. (Figura 121).

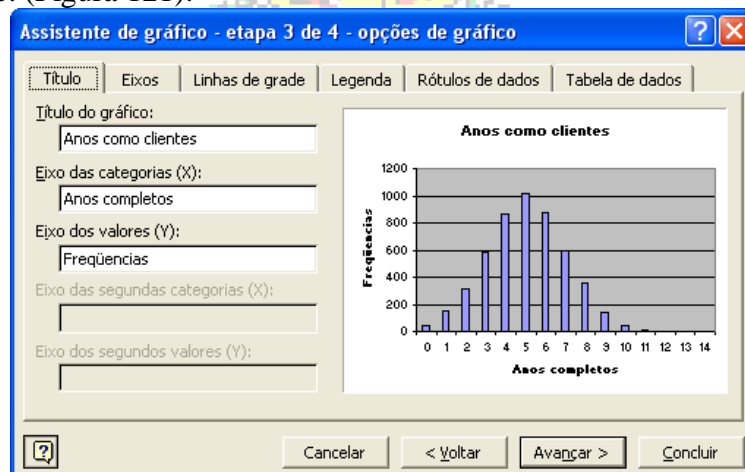


Figura 121 - Assistente de gráfico 3a etapa - gráfico de variável quantitativa – Título
O gráfico resultante está na Figura 122.



Figura 122 - Histograma de Anos como clientes

Foi dito anteriormente 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 123)

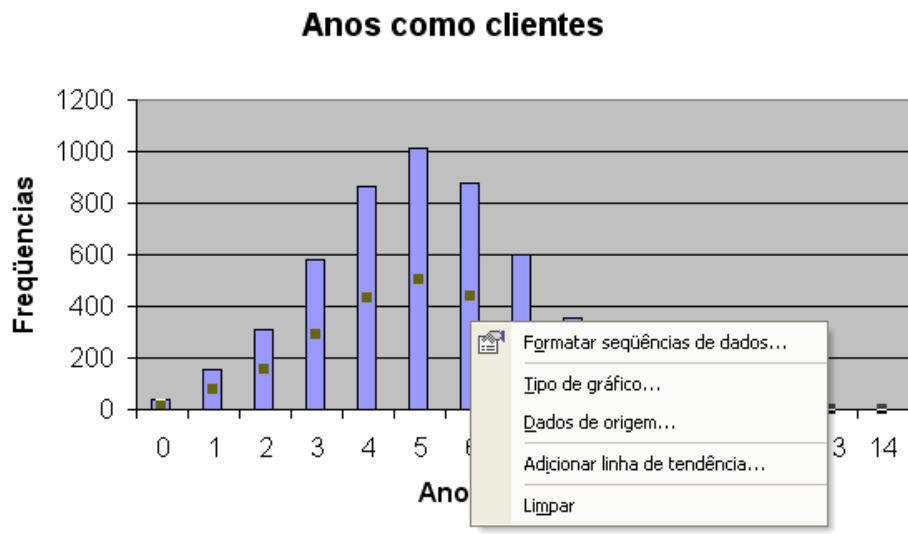


Figura 123 - Histograma de Anos - opções das colunas

Escolhendo “Formatar seqüência de dados”, chegamos à Figura 124.

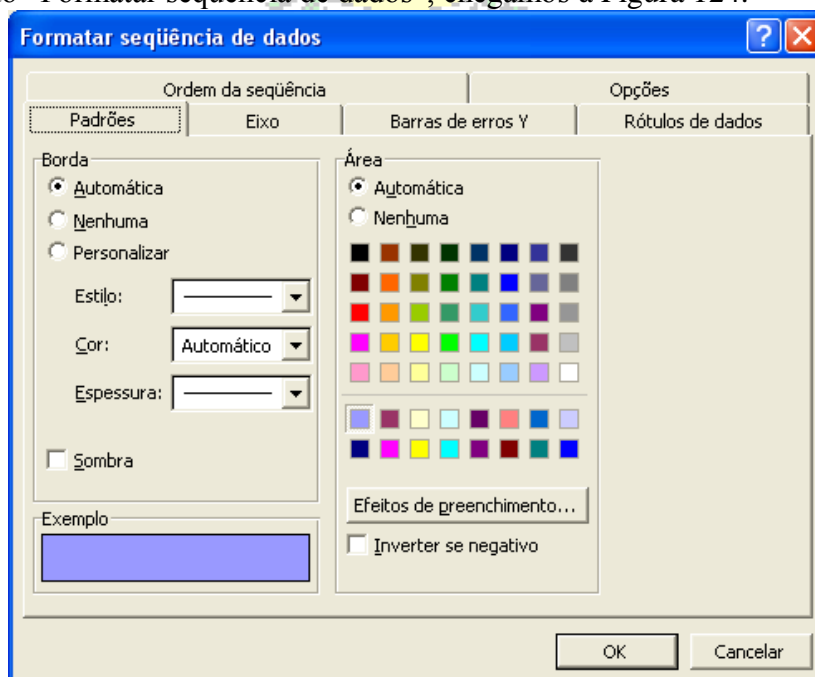


Figura 124 - Opções de formatar seqüência de dados

Dentre as várias possibilidades devemos selecionar “Opções”, mostrada na Figura 125.

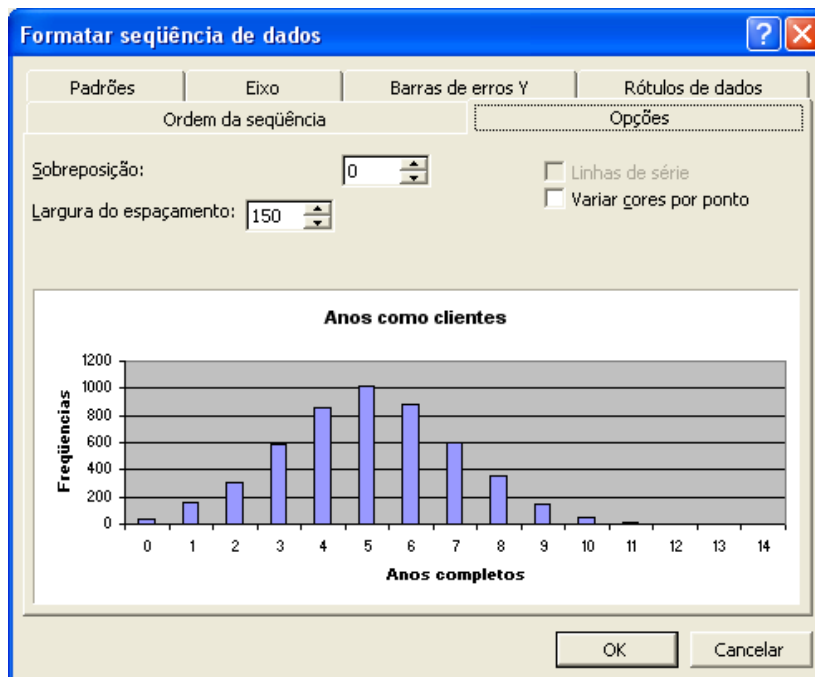


Figura 125 - Opções de Formatar seqüência de dados

Reduzindo o campo “Largura do espaçamento” a zero, o histograma correto é mostrado na Figura 126.

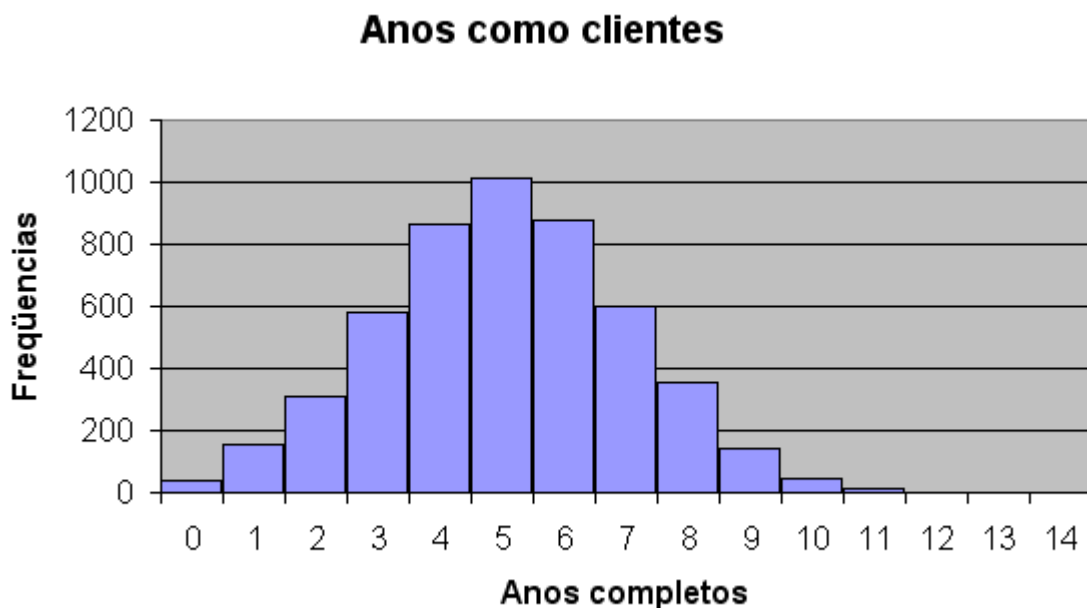


Figura 126 - Histograma de Anos completos como cliente

Percebe-se claramente que os anos estão concentrados em torno de 5, diminuindo as frequências progressivamente, à medida que se afastam de 5.

5.2 - Procedimentos para variáveis quantitativas contínuas

Se a variável de interesse for quantitativa contínua, como Renda no arquivo Celular01.xlsx, será necessário construir uma tabela de frequências agrupada em classes, ou seja agrupar os dados. Como os valores de uma variável contínua repetem-se pouco (ou não se repetem), teríamos uma tabela dinâmica imensa (com talvez centenas de linha no caso da variável Renda), se isso não fosse feito.

Vamos observar os passos necessários:

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 W2 e o máximo na célula W3, . Selecione esta célula com o cursor. Podemos inserir as funções ou escrevê-las diretamente nas células:

= MÍNIMO(G2:G5001)

= MÁXIMO(G2:G5001)

Veja a Figura 127.

	V	W	X	Y
1	Renda			
2	Mínimo	=MÍNIMO(G2:G5001)		
3	Máximo			

	V	W	X	Y
1	Renda			
2	Mínimo	1,003		
3	Máximo	=MÁXIMO(G2:G5001)		

Figura 127 - Mínimo e máximo para a variável Renda

Para calcular o intervalo basta colocar a fórmula respectiva em uma célula, W4 por exemplo. Veja a Figura 128.

	V	W	X
1	Renda		
2	Mínimo	1,003	
3	Máximo	43,878	
4	Intervalo	=W3-W2	

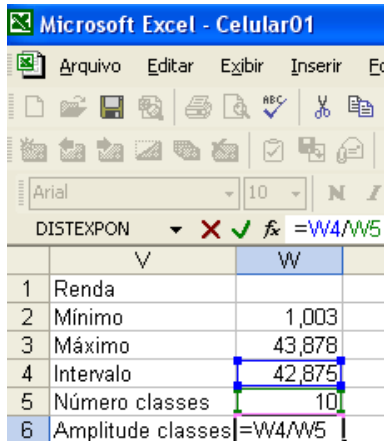
Figura 128 - Intervalo para Renda

Observa-se, conforme visto em aulas anteriores usando o Auto-filtro, que a Renda varia de 1,003 a 43,878 salários mínimos, resultando em um intervalo de 42,875 salários mínimos, que precisa ser dividido em um número conveniente de classes.

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

Usualmente definimos o número de classes calculando a raiz quadrada do tamanho do conjunto. Na variável Renda há 4996 observações, extraindo a raiz quadrada (a função RAIZ, na

categoria "Matemática e trigonometria" do Excel), obtemos 70,68. Poderíamos utilizar 71 classes, mas este valor seria elevado (criaríamos muitas categorias, o que talvez prejudicasse nosso objetivo de resumir o conjunto de dados). Por outro lado, usar poucas classes pode resumir "demais" o conjunto, forçando o agrupamento. Arbitrariamente usaremos 10 classes. Com isso a amplitude das classes seria igual a 42,875 (intervalo) dividido por 10, resultando que cada classe teria uma amplitude de 4,2875. Veja a Figura 129.

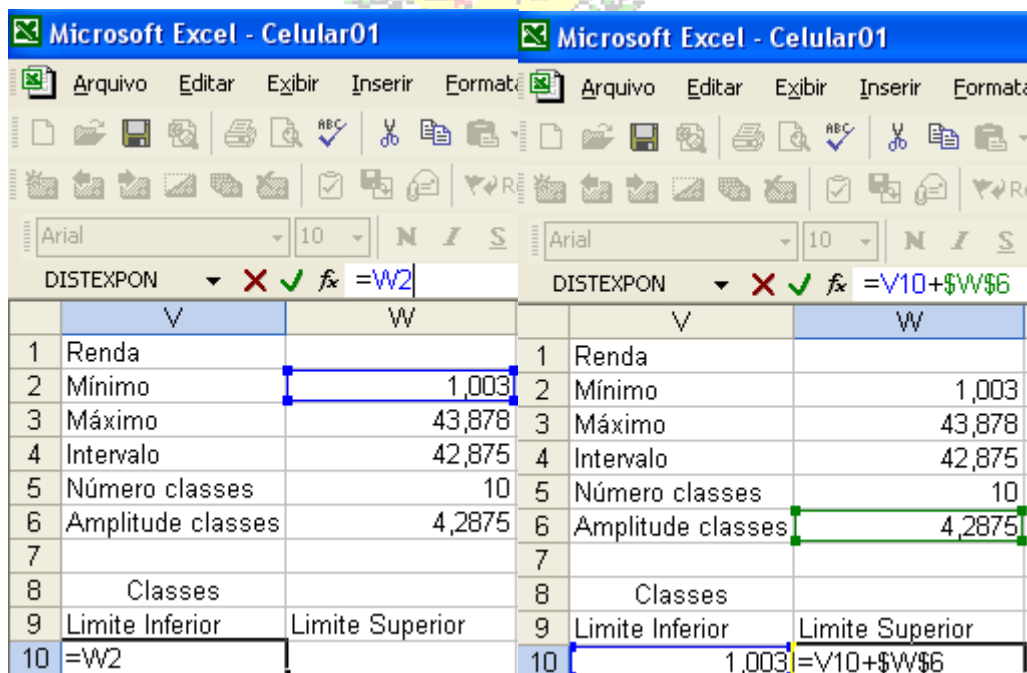


	V	W
1	Renda	
2	Mínimo	1,003
3	Máximo	43,878
4	Intervalo	42,875
5	Número classes	10
6	Amplitude classes	=W4/W5

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 1,003 (1, por exemplo), e a amplitude garanta que o valor máximo será incluído no conjunto. Escolhendo uma amplitude de 4,2875, e um valor inicial igual ao mínimo do conjunto, tendo em mente que a tabela terá 10 classes, o resultado poderá ser calculado em células apropriadas, como mostrado a partir da Figura 130.

Figura 129 - Amplitude de classes de Renda



	V	W		V	W
1	Renda		1	Renda	
2	Mínimo	1,003	2	Mínimo	1,003
3	Máximo	43,878	3	Máximo	43,878
4	Intervalo	42,875	4	Intervalo	42,875
5	Número classes	10	5	Número classes	10
6	Amplitude classes	4,2875	6	Amplitude classes	4,2875
7			7		
8	Classes		8	Classes	
9	Limite Inferior	Limite Superior	9	Limite Inferior	Limite Superior
10	=W2		10	1,003	=V10+\$W\$6

Figura 130 – Cálculo dos limites inferior e do limite superior da primeira classe de Renda

O limite inferior da primeira classe será o conteúdo da célula W2 (mínimo de Renda), e o limite superior será o limite inferior (célula W10) mais o conteúdo da célula W6 (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 131.

	V	W		V	W
1	Renda		1	Renda	
2	Mínimo	1,003	2	Mínimo	1,003
3	Máximo	43,878	3	Máximo	43,878
4	Intervalo	42,875	4	Intervalo	42,875
5	Número classes	10	5	Número classes	10
6	Amplitude classes	4,2875	6	Amplitude classes	4,2875
7			7		
8	Classes		8	Classes	
9	Limite Inferior	Limite Superior	9	Limite Inferior	Limite Superior
10	1,003	5,2905	10	1,003	5,2905
11	=W10		11	5,2905	=V11+\$W\$6

Figura 131 – 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 V11 e W11 até o fim da tabela e completar as 10 classes. Veja a Figura 132.

	V	W
1	Renda	
2	Mínimo	1,003
3	Máximo	43,878
4	Intervalo	42,875
5	Número classes	10
6	Amplitude classes	4,2875
7		
8	Classes	
9	Limite Inferior	Limite Superior
10	1,003	5,2905
11	5,2905	9,578
12	9,578	13,8655
13	13,8655	18,153
14	18,153	22,4405
15	22,4405	26,728
16	26,728	31,0155
17	31,0155	35,303
18	35,303	39,5905
19	39,5905	43,878

Figura 132 - Classes de Renda

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

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

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

Esses procedimentos são mostrados nas Figuras 133 a 135.

	V	W	X	Y	Z
1	Renda				
2	Mínimo	1,003			
3	Máximo	43,878			
4	Intervalo	42,875			
5	Número classes	10			
6	Amplitude classes	4,2875			
7					
8	Classes				
9	Limite Inferior	Limite Superior	Ponto médio	Freqüência	%
10	1,003	5,2905	$= (V10+W10)/2$		
11	5,2905	9,578			
12	9,578	13,8655			
13	13,8655	18,153			
14	18,153	22,4405			
15	22,4405	26,728			
16	26,728	31,0155			
17	31,0155	35,303			
18	35,303	39,5905			
19	39,5905	43,878			

Figura 133 - Ponto médio da primeira classe de Renda

Basta arrastar a fórmula da célula X10 até a X19 para obter os pontos médios das classes.

	V	W	X	Y	Z	AA	AB	AC	AD	AE
8	Classes									
9	Limite Inferior	Limite Superior	Ponto médio	Freqüência	%					
10	1,003	5,2905	3,14675	$= \text{CONT.SE}(\$G\$2:\$G\$5001;"<"\&W10) - \text{CONT.SE}(\$G\$2:\$G\$5001;"<"\&V10)$						
11	5,2905	9,578								
12	9,578	13,8655								
13	13,8655	18,153								
14	18,153	22,4405								
15	22,4405	26,728								
16	26,728	31,0155								
17	31,0155	35,303								
18	35,303	39,5905								
19	39,5905	43,878								

Figura 134 - Obtenção da freqüência da primeira classe de Renda

Observe que contam-se as ocorrências menores do que o limite superior da classe, subtraídas da contagem das ocorrências menores do que o limite inferior, mantendo o intervalo G2 a G5001 (valores de Renda) como referência absoluta, para permitir arrastar a fórmula até a célula Y19.

	V	W	X	Y	Z	AA	AB	AC	AD	AE
8	Classes									
9	Limite Inferior	Limite Superior	Ponto médio	Frequência	%					
10	1,003	5,2905	3,14675	2959						
11	5,2905	9,578	7,43425	1196						
12	9,578	13,8655	11,72175	493						
13	13,8655	18,153	16,00925	183						
14	18,153	22,4405	20,29675	97						
15	22,4405	26,728	24,58425	42						
16	26,728	31,0155	28,87175	12						
17	31,0155	35,303	33,15925	7						
18	35,303	39,5905	37,44675	5						
19	39,5905	43,878	41,73425	=CONT.SE(\$G\$2:\$G\$5001;"<="&W19)-CONT.SE(\$G\$2:\$G\$5001;"<"&V19)						

Figura 135 - 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.

Podemos também obter frequências e percentuais acumulados para a tabela de Renda. Basta remover o segundo membro das fórmulas mostradas nas Figuras 134 e 135, tomando os mesmos cuidados com o limite superior da última classe, como mostrado nas Figuras 136 e 137.

	V	W	X	Y	Z	AA	AB
8	Classes						
9	Limite Inferior	Limite Superior	Ponto médio	Frequência	%	Freq.Acumulada	% acumulado
10	1,003	5,2905	3,14675	2959	59,23%	=CONT.SE(G2:G5001;"<"&W10)	
11	5,2905	9,578	7,43425	1196	23,94%		
12	9,578	13,8655	11,72175	493	9,87%		
13	13,8655	18,153	16,00925	183	3,66%		
14	18,153	22,4405	20,29675	97	1,94%		
15	22,4405	26,728	24,58425	42	0,84%		
16	26,728	31,0155	28,87175	12	0,24%		
17	31,0155	35,303	33,15925	7	0,14%		
18	35,303	39,5905	37,44675	5	0,10%		
19	39,5905	43,878	41,73425	2	0,04%		
20			Total	4996	100,00%		

Figura 136 - Obtenção da frequência acumulada da primeira classe de Renda

	V	W	X	Y	Z	AA	AB	AC
8	Classes							
9	Limite Inferior	Limite Superior	Ponto médio	Freqüência	%	Freq.Acumulada	% acumulado	
10	1,003	5,2905	3,14675	2959	59,23%	2959		
11	5,2905	9,578	7,43425	1196	23,94%	4155		
12	9,578	13,8655	11,72175	493	9,87%	4648		
13	13,8655	18,153	16,00925	183	3,66%	4831		
14	18,153	22,4405	20,29675	97	1,94%	4928		
15	22,4405	26,728	24,58425	42	0,84%	4970		
16	26,728	31,0155	28,87175	12	0,24%	4982		
17	31,0155	35,303	33,15925	7	0,14%	4989		
18	35,303	39,5905	37,44675	5	0,10%	4994		
19	39,5905	43,878	41,73425	2	0,04%	=CONT.SE(\$G\$2:\$G\$5001;"<="&W19)		
20			Total	4996	100,00%			

Figura 137 - Obtenção da freqüência acumulada da primeira classe de Renda

Após todas as operações, incluindo a obtenção dos percentuais a tabela agrupada em classes da Renda resulta na Figura 138.

	V	W	X	Y	Z	AA	AB
8	Classes						
9	Limite Inferior	Limite Superior	Ponto médio	Freqüência	%	Freq.Acumulada	% acumulado
10	1,003	5,2905	3,14675	2959	59,23%	2959	59,23%
11	5,2905	9,578	7,43425	1196	23,94%	4155	83,17%
12	9,578	13,8655	11,72175	493	9,87%	4648	93,03%
13	13,8655	18,153	16,00925	183	3,66%	4831	96,70%
14	18,153	22,4405	20,29675	97	1,94%	4928	98,64%
15	22,4405	26,728	24,58425	42	0,84%	4970	99,48%
16	26,728	31,0155	28,87175	12	0,24%	4982	99,72%
17	31,0155	35,303	33,15925	7	0,14%	4989	99,86%
18	35,303	39,5905	37,44675	5	0,10%	4994	99,96%
19	39,5905	43,878	41,73425	2	0,04%	4996	100,00%
20			Total	4996	100,00%		

Figura 138 - Renda agrupada em classes: freqüências absolutas, percentuais, freqüências acumuladas e percentuais acumulados

A maioria absoluta (na realidade, mais, pois são 59,23%) dos clientes da operadora têm renda de até 5,29 salários mínimos. Apenas 7% dos clientes têm renda acima de 13,86 salários mínimos, o que pode ser importante para a empresa direcionar seu marketing. O importante, também, é que a variável foi resumida em uma tabela que facilita sua análise.

5.2.1 – Histograma para variáveis quantitativas contínuas

Podemos fazer um raciocínio análogo para uma variável quantitativa contínua, representada em uma tabela agrupada em classes, como na Figura 138: acionar o assistente gráfico, selecionar os valores (Figura 139), selecionar os rótulos do eixo X (Figura 140).

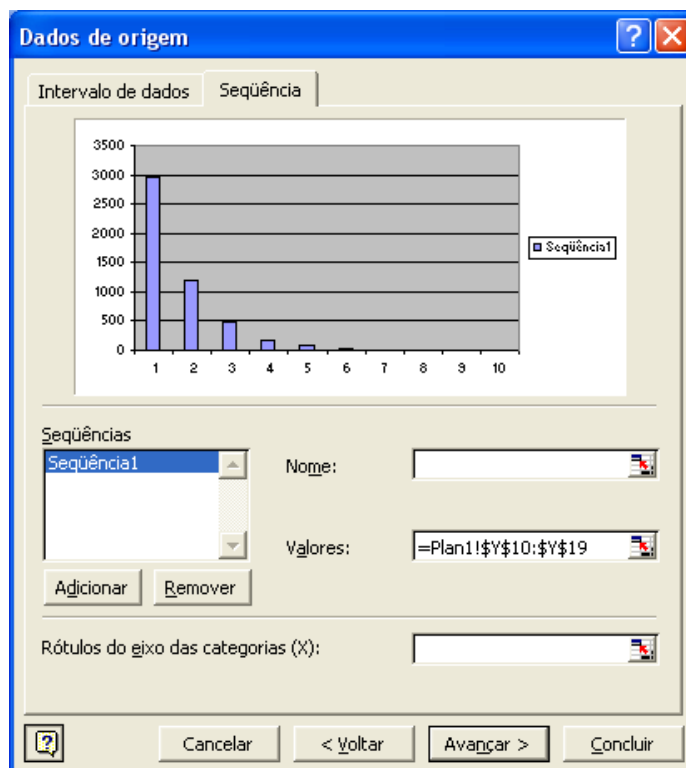


Figura 139 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – valores adicionados

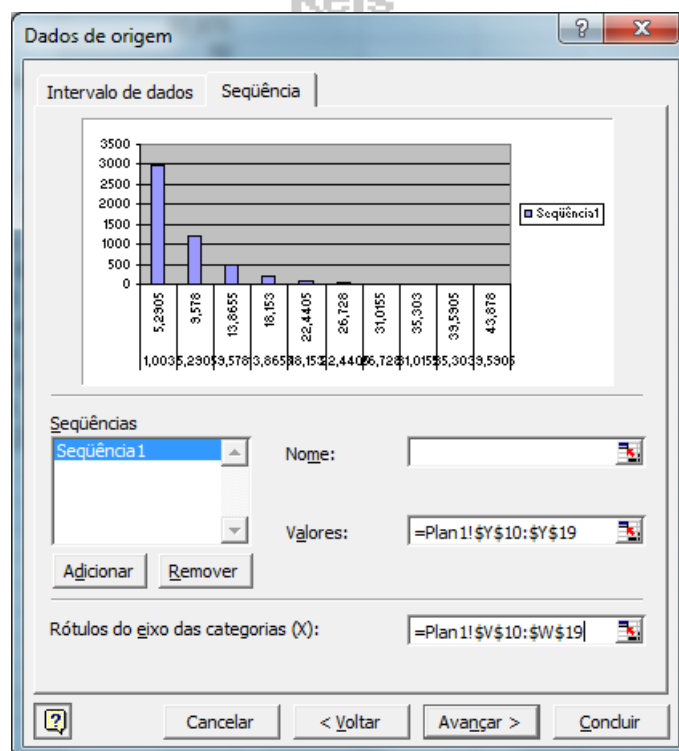


Figura 140 - Assistente de gráfico 2a etapa - gráfico de variável quantitativa – seleção de rótulos do eixo X

Escolhermos pôr nos rótulos do eixo das categorias os limites das classes, mas poderíamos escolher os pontos médios que são os representantes das classes. Veja a Figura 141, já com os espaçamentos entre as colunas iguais a zero e a área de plotagem com a cor modificada.




Figura 141 - Histograma da Renda dos clientes em salários mínimos com limites das classes

Percebe-se que a maioria esmagadora dos clientes tem renda até 5,2905 salários mínimos, e que há muito poucos clientes com renda acima de 22,4405 salários mínimos. São as mesmas conclusões da tabela da Figura 138, mas a apreensão da informação é mais rápida através do gráfico.

5.3 – Medidas de síntese

Uma das maiores vantagens do Excel (seja qual versão for) é a facilidade de cálculo das medidas de síntese. Basta inserir a função, fornecer os argumentos apropriados e obter os resultados. É possível calcular as medidas de síntese para uma variável quantitativa individualmente, ou em função de uma ou mais variáveis qualitativas (através de uma tabela dinâmica ou não). Vamos começar com o cálculo individual. Para tanto, no arquivo Celular01.xlsx, devemos selecionar a planilha Plan1, onde estão os dados. Repare que estes dados já estão classificados em função da variável Tipo (ver seção 2), o que nos será útil em breve.

5.3.1 – 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 127). Ou, podemos usar os menus do Excel 2003: precisamos procurar pelo ícone "Inserir Função" (), ver Figura 142.

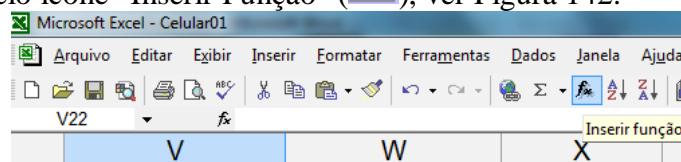
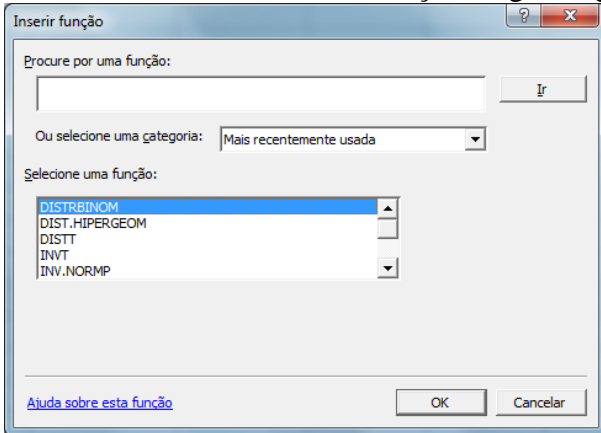


Figura 142 - Fórmulas no Excel 2007

Pressionando Inserir Função surge a Figura 143.



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 144), 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.

Figura 143 - Funções disponíveis no Excel

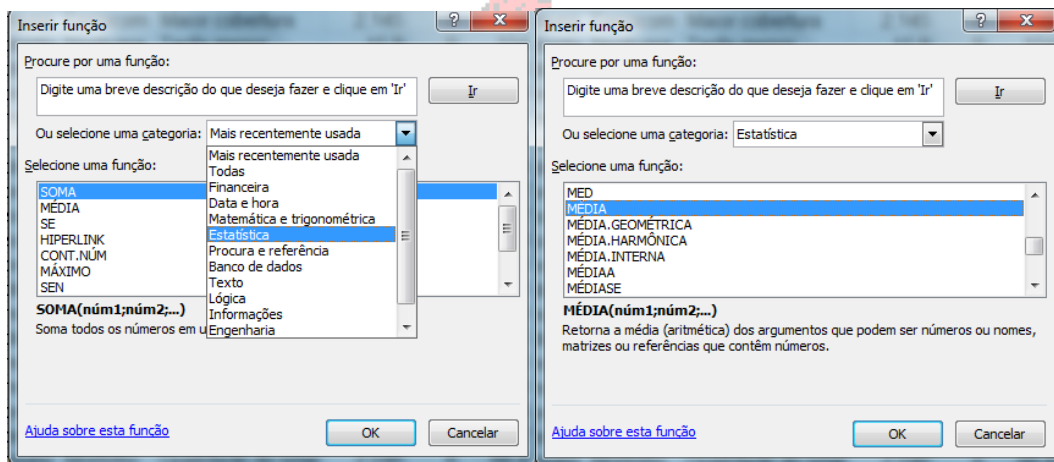


Figura 144 - Funções estatísticas no Excel

Na tela à direita da Figura 144 podemos escolher calcular a média.

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 145.

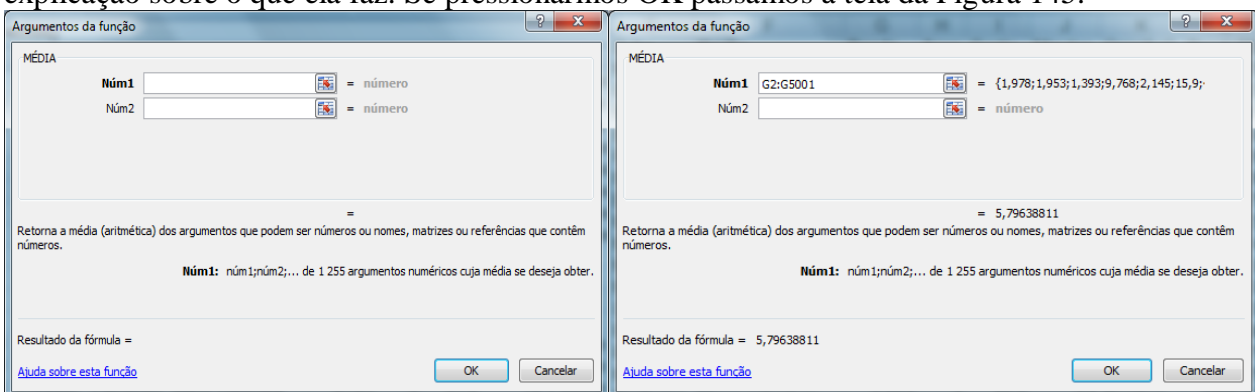


Figura 145 - Argumentos da função Média (para Renda)

Os dados de Renda estão nas células G2 a G5001¹⁰ da planilha onde estão os dados. Então, ou selecionamos os dados diretamente (pressionando à seta vermelha ao lado dos campos “Núm1” ou “Núm2” e escolhendo as células com o mouse) ou escrevemos o intervalo G2:G5001. Os dois pontos indicam para o Excel que todas as células do intervalo precisam ser consideradas no cálculo,

¹⁰ A célula G1 é excluída por conter o nome da variável.

mas a função desconsidera células vazias. Antes mesmo de pressionar OK o Excel já apresenta o resultado: 5,796 salários mínimos (cerca de R\$ 2956, em agosto de 2010). Se repetirmos o processo para as outras medidas, completamos os cálculos, os procedimentos estão nas Figuras 146 a 149.

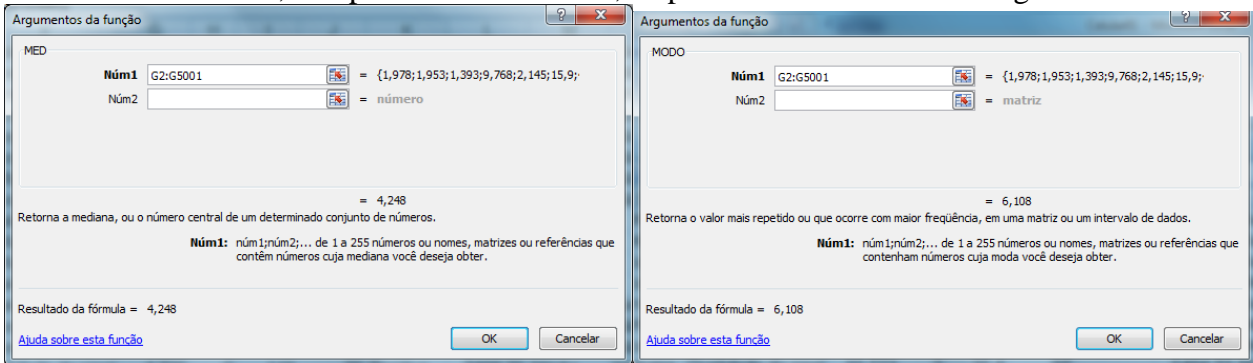
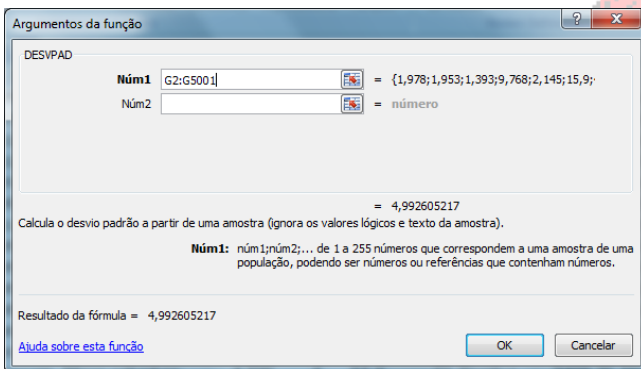


Figura 146 - Argumentos das funções Mediana e Moda (Renda)



É importante ressaltar que o arquivo Celular01.xls 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 147 - Argumentos da função desvio padrão (Renda)

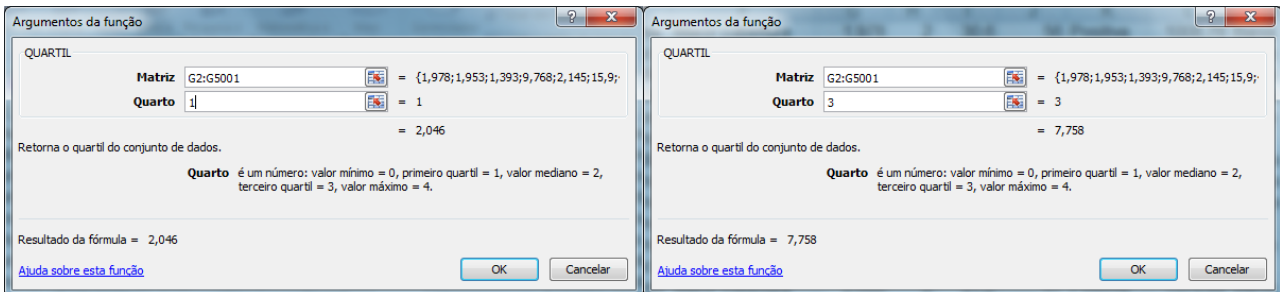


Figura 148 - Argumentos das funções Quartil Inferior e Superior (Renda)

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.

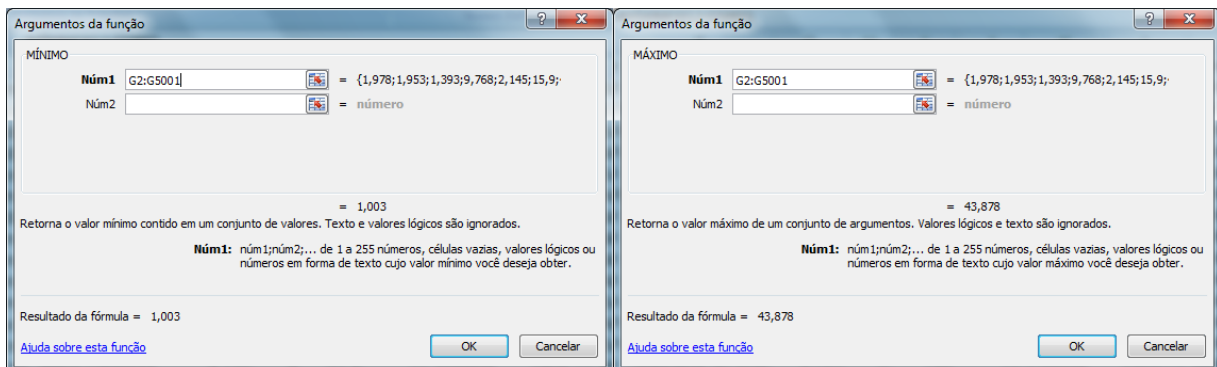


Figura 149 - Argumentos das funções Mínimo e Máximo (Renda)

E os resultados completos podem ser vistos na Figura 150.

	Z	AA
1	Medidas de Renda	
2	Média	5,796388
3	Mediana	4,248
4	Moda	6,108
5	Desvio padrão	4,992605
6	Quartil inferior	2,046
7	Quartil superior	7,758
8	Mínimo	1,003
9	Máximo	43,878

Figura 150 - 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 150: coeficiente de variação percentual, diferenças entre os quartis e a mediana, e os valores que definem os valores discrepantes. Nas Figuras 151 a 153 apresentamos as operações:

	Z	AA	AB	AC	AD
1	Medidas de Renda				
2	Média	5,796388	CV%	$=(AA5/AA2)*100$	
3	Mediana	4,248			
4	Moda	6,108			
5	Desvio padrão	4,992605			
6	Quartil inferior	2,046			
7	Quartil superior	7,758			
8	Mínimo	1,003			
9	Máximo	43,878			

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 151 - Cálculo do CV%

	Z	AA	AB	AC		Z	AA	AB	AC	
1	Medidas de Renda					1	Medidas de Renda			
2	Média	5,796388	CV%	86,13304		2	Média	5,796388	CV%	86,13304
3	Mediana	4,248	Qs-Md	$=AA7-AA3$		3	Mediana	4,248	Qs-Md	3,51
4	Moda	6,108				4	Moda	6,108	Md-Qi	$=AA3-AA6$
5	Desvio padrão	4,992605				5	Desvio padrão	4,992605		
6	Quartil inferior	2,046				6	Quartil inferior	2,046		
7	Quartil superior	7,758				7	Quartil superior	7,758		
8	Mínimo	1,003				8	Mínimo	1,003		
9	Máximo	43,878				9	Máximo	43,878		

Figura 152 - 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 a variável pode ser considerado simétrico em relação à tendência central (além disso, média e mediana devem ser próximas também).

	Z	AA	AB	AC	AD		Z	AA	AB	AC	AD	
1	Medidas de Renda						1	Medidas de Renda				
2	Média	5,796388	CV%	86,13304		2	Média	5,796388	CV%	86,13304		
3	Mediana	4,248	Qs-Md	3,51		3	Mediana	4,248	Qs-Md	3,51		
4	Moda	6,108	Md-Qi	2,202		4	Moda	6,108	Md-Qi	2,202		
5	Desvio padrão	4,992605	$Qs+1,5x(Qs-Qi)$	$=AA7+1,5*(AA7-AA6)$		5	Desvio padrão	4,992605	$Qs+1,5x(Qs-Qi)$	16,326		
6	Quartil inferior	2,046				6	Quartil inferior	2,046	$Qi-1,5x(Qs-Qi)$	$=AA6-1,5*(AA7-AA6)$		
7	Quartil superior	7,758				7	Quartil superior	7,758				
8	Mínimo	1,003				8	Mínimo	1,003				
9	Máximo	43,878				9	Máximo	43,878				

Figura 153 - Cálculo dos limites para valores discrepantes

Valores discrepantes inferiores são todos aqueles abaixo do limite $Qi - 1,5 \times (Qs - Qi)$. E discrepantes superiores todos acima de $Qs + 1,5 \times (Qs - Qi)$. Os resultados completos estão na Figura 154.

	Z	AA	AB	AC
1	Medidas de Renda			
2	Média	5,796388	CV%	86,13304
3	Mediana	4,248	Qs-Md	3,51
4	Moda	6,108	Md-Qi	2,202
5	Desvio padrão	4,992605	Qs+1,5x(Qs-Qi)	16,326
6	Quartil inferior	2,046	Qi-1,5x(Qs-Qi)	-6,522
7	Quartil superior	7,758		
8	Mínimo	1,003		
9	Máximo	43,878		

Figura 154 - Medidas de síntese para Renda (completas)

Tendência Central:

- o valor típico de Renda oscila entre 4,248 (mediana - 50% dos clientes abaixo e 50% acima deste valor) e 5,796 salários mínimos (média);
- ou seja, entre R\$2166,48 e R\$2955,96 (valores de agosto de 2010);
- 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 duas vezes maior do que a média do Brasil¹¹.
- a moda vale 6,108 salários mínimos, mas para variáveis quantitativas contínuas esta medida não tem muito significado, pois os valores se repetem pouco¹².

Dispersão:

- a variação total da renda é de 1,003 (mínimo) a 43,878 (máximo) salários mínimos (R\$511,53 a R\$22377,78 – valores de agosto de 2010), a maior renda é 43 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 86,13% da média;
- 25% dos clientes têm renda de até 2,046 (quartil inferior) salários mínimos (R\$1043,56), e 25% têm renda acima de 7,758 (quartil superior) salários mínimos (R\$3956,58) – em agosto de 2010;
- como consequência da afirmação anterior, 50% dos clientes têm renda entre 2,046 e 7,758 salários mínimos.

Assimetria

- média e mediana são diferentes, e como a média é maior do que a mediana a assimetria é à direita;
- a conclusão acima é confirmada pela diferença entre quartil superior e mediana (3,51) que é maior do que a entre a mediana e o quartil inferior (2,202).

Valores discrepantes

- não há discrepantes inferiores, pois precisaria haver renda inferior -6,522 salários mínimos;
- há discrepantes superiores, pois há valores acima¹³ de 16,326 salários mínimos (R\$8326,26 em agosto de 2010).

5.3.2 – Medidas de síntese para uma variável quantitativa em função de uma qualitativa – por classificação

Uma das ferramentas mais úteis de análise exploratória de dados de variáveis quantitativas é o “breakdown”: consiste em calcular medidas de síntese de uma variável quantitativa em função dos valores de uma ou mais variáveis qualitativas (que podem ser chamadas de variáveis de agrupamento ou independentes). Com base nos resultados encontrados é possível avaliar se as variáveis apresentam algum relacionamento:

- da mesma forma que para as variáveis qualitativas devem-se comparar os valores totais com os

¹¹ É FUNDAMENTAL ter algum padrão para comparação, o que foi possível neste caso.

¹² Usando a função CONT.SE descobre-se que a moda ocorreu apenas 31 vezes em um total de 5000 observações.

¹³ Usando a função CONT.SE verifica-se que há 224 clientes com renda acima deste valor, sendo portanto valores discrepantes.

obtidos para cada valor das variáveis e agrupamento;

- se houver diferenças significativas entre as medidas totais (com todos os dados) e as para cada grupo formado em função dos valores das variáveis de agrupamento há evidência de relacionamento entre as variáveis.

Imagine que há interesse em avaliar se os clientes de Pós-pago e Pré-pago têm perfil de renda diferente. Neste caso há duas variáveis: Tipo, que será a independente (ou de agrupamento) e Renda, que será a dependente. Na seção 5.3.1 calculamos as medidas de síntese de Renda, sem segmentação (as medidas totais). Agora precisamos calcular as mesmas medidas de síntese para os clientes de Pós-pago e Pré-pago. Isso exigirá a classificação dos dados em função da variável Tipo, o que já foi feito na seção 2 (ver Figuras 17 a 19). Se inspecionarmos o conjunto classificado vamos observar: as células vazias de Tipo vão da célula A2 até a A8; as células de clientes de Pós-pago vão da célula A9 até a A2110; as células de clientes de Pré-pago vão da célula A2111 até a A5001. Não vamos calcular medidas para as células vazias de Tipo. Temos que repetir o procedimento da seção 5.3.1 para as células G9 a G2110 (medidas de síntese de Renda para os clientes de Pós-pago) e das células G2111 a G5001 (medidas de síntese de Renda para os clientes de Pré-pago). Basta repetir os passos das Figuras 146 a 149 trocando o intervalo de células lá utilizado pelos mencionados acima, resultando na Figura 155:

	AE	AF	AG	AH		AJ	AK	AL	AM
1	Medidas de Renda Pós-pago				1	Medidas de Renda Pré-pago			
2	Média	8,3839	CV%	68,6016713	2	Média	3,918023	CV%	83,16532
3	Mediana	6,9	Qs-Md	4,197	3	Mediana	2,811	Qs-Md	2,331
4	Moda	5,001	Md-Qi	2,568	4	Moda	2,934	Md-Qi	1,127
5	Desvio padrão	5,751495	Qs+1,5x(Qs-Qi)	21,2445	5	Desvio padrão	3,258437	Qs+1,5x(Qs-Qi)	10,329
6	Quartil inferior	4,332	Qi-1,5x(Qs-Qi)	-5,8155	6	Quartil inferior	1,684	Qi-1,5x(Qs-Qi)	-3,503
7	Quartil superior	11,097			7	Quartil superior	5,142		
8	Mínimo	1,032			8	Mínimo	1,003		
9	Máximo	43,878			9	Máximo	35,796		

	Z	AA	AB	AC
1	Medidas de Renda			
2	Média	5,796388	CV%	86,13304
3	Mediana	4,248	Qs-Md	3,51
4	Moda	6,108	Md-Qi	2,202
5	Desvio padrão	4,992605	Qs+1,5x(Qs-Qi)	16,326
6	Quartil inferior	2,046	Qi-1,5x(Qs-Qi)	-6,522
7	Quartil superior	7,758		
8	Mínimo	1,003		
9	Máximo	43,878		

Figura 155 - Breakdown de Renda por Tipo - a partir do arquivo classificado

Precisamos repetir a descrição da seção 5.3.1, mas comparando as medidas por Tipo e em relação às totais. Vamos analisar apenas por salários mínimos, deixando ao leitor a tarefa de conversão dos valores para reais.

Tendência Central:

- o valor típico geral de Renda oscila entre 4,248 (mediana) e 5,796 salários mínimos (média);
- tanto a média quanto a mediana dos clientes do Pós-pago são maiores do que o geral, e também maiores do que as do Pré-pago – 8,3839 e 6,9 contra 3,918 e 2,811 respectivamente;
- conclui-se que os clientes de Pós-pago têm maior tendência central de Renda, com diferença de 4,46 salários mínimos na média e 4,089 na mediana, respectivamente, em relação ao Pré-pago;

Dispersão:

- a variação total da renda é de 1,003 (mínimo) a 43,878 (máximo), sendo que é mais acentuada no Pós-pago (1,032 a 43,878- diferença de 43 vezes) do que no Pré-pago (1,003 a 35,796 – diferença de “apenas” 35 vezes), mas em ambos a dispersão é grande;
- a grande dispersão é confirmada pelo alto valor do coeficiente de variação percentual, indicando

que o desvio padrão representa 86,13% da média no total, mas é menor no Pós-pago (68,60%) do que no Pré-pago (83,16%);

- no total, 25% dos clientes têm renda de até 2,046 (quartil inferior) salários mínimos, e 25% têm renda acima de 7,758 (quartil superior) salários mínimos;

- como consequência da afirmação anterior, 50% do total de clientes têm renda entre 2,046 e 7,758 salários mínimos;

- os quartis do Pós-pago são ambos maiores do os do Pré-pago (e do total de clientes), 4,332 e 11,097 salários mínimos contra 1,684 e 5,142 respectivamente – apenas 25% dos clientes de Pré-pago têm renda acima de 5,142 salários mínimos (o quartil *superior* do Pré-pago nem sequer atinge a *mediana* do Pós-pago que vale 6,9);

- 50% dos clientes do Pós-pago tem renda entre 4,332 e 11,097 salários mínimos, enquanto que no Pré-pago, 50% têm renda entre 1,684 e 5,142 salários mínimos;

Assimetria

- média e mediana são diferentes no total de clientes, e como a média é maior do que a mediana a assimetria é à direita;

- a conclusão acima é confirmada pela diferença entre quartil superior e mediana (3,51) que é maior do que a entre a mediana e o quartil inferior (2,202).

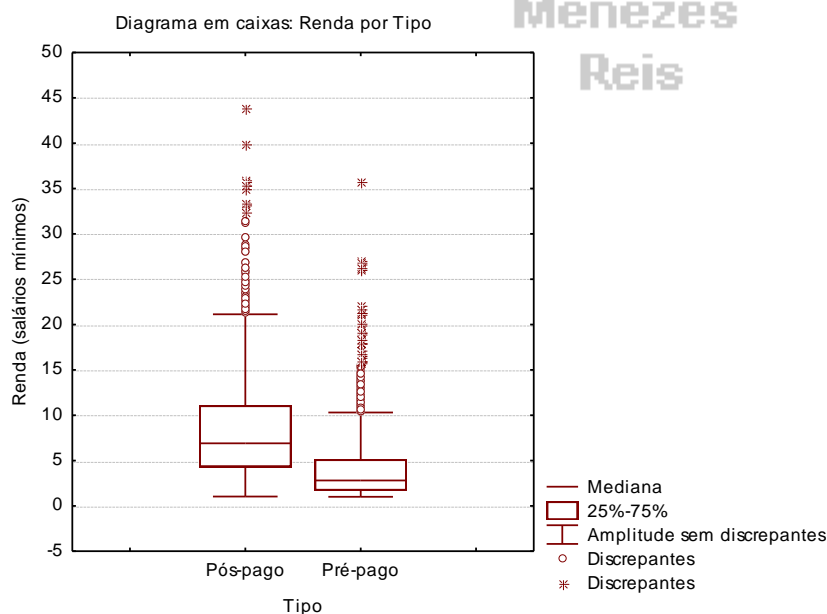
- média e mediana também são diferentes nos dois Tipos, constatando-se também pelas diferenças entre quartis e mediana;

Valores discrepantes

- não há discrepantes inferiores, pois precisaria haver renda inferior -6,522 salários mínimos no total de clientes, inferior a -5,815 no Pré-pago e -3,503 no Pós-pago;

- há discrepantes superiores, pois há valores acima de 16,326 salários mínimos no total de clientes, 21,4445 no Pós-pago e 10,329 no Pré-pago¹⁴.

As diferenças entre as Rendas dos clientes de Pós-pago e Pré-pago podem ser detectadas em um diagrama em caixas múltiplo (Figura 156):



O valor típico (mediana) de Pós-pago é maior do que o do Pré-pago, e o quartil superior do Pré-pago é pouco maior do que o do Pós-pago: ou seja, cerca de 75% dos clientes de Pós-pago têm renda correspondente aos 25% mais abastados clientes de Pré-pago.

Há assimetria em ambos, e a dispersão aparenta ser maior em Pós-pago ($Q_s - Q_i$) maior do que no Pré-pago.

Há valores discrepantes superiores nos dois grupos.

Figura 156 - Diagrama em caixas de Renda por Tipo

O procedimento poderia ser estendido para os casos em que mais de uma variável de agrupamento, por exemplo, poderíamos fazer um breakdown de Renda por Tipo e por Opinião.

¹⁴ Usando a função CONT.SE verifica-se que há 224 clientes com renda acima de 16,326 salários mínimos no total, 78 com renda acima de 21,4445 no Pós-pago e 133 com renda acima de 10,329 no Pré-pago.

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

É possível realizar o breakdown descrito na seção 5.3.2 usando tabelas dinâmicas, o que evita a necessidade de classificação prévia dos dados. Mas, nas tabelas dinâmicas não há todas as medidas disponíveis no repertório de funções estatísticas do Excel: há apenas contagem (e a possibilidade de apresentar os resultados como percentuais), mínimo, máximo, média, variância (amostral e populacional) e desvio padrão (amostral e populacional). O procedimento é semelhante ao caso de duas variáveis qualitativas, mas é preciso tomar alguns cuidados.

O início do processo é semelhante ao visto na seção 4.1: menu Dados, opção Relatório de Tabela e Gráfico Dinâmicos. Posteriormente, precisamos definir onde estão os dados: no nosso caso, na Plan1, células A1 a N5001 do arquivo Celular01.xls, como mostrado na Figura 49. Seguindo os procedimentos mostrados nas Figuras 50 e 51 chegamos à Figura 157, indicando que colocaremos os resultados em uma planilha existente:

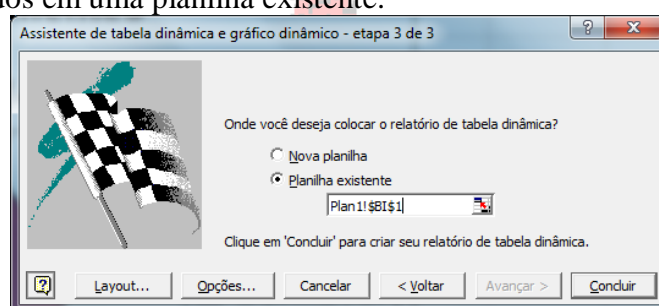


Figura 157 - Etapa 3 da tabela dinâmica – Renda por Tipo

Se pressionarmos Layout na Figura 157 vamos arrastar a variável Tipo para o campo LINHA, e Renda para o campo DADOS, NÃO arraste nenhuma outra variável para o campo colunas, resultando na Figura 158

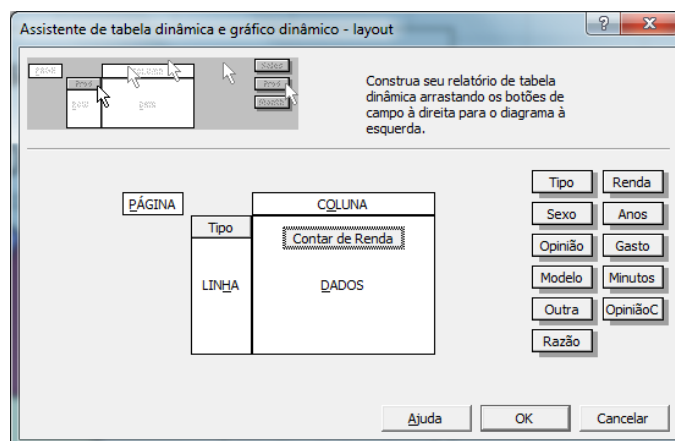


Figura 158 - Layout da tabela dinâmica Renda x Tipo 1a parte

No caso acima a tabela apenas contaria quantos clientes há em cada Tipo de serviço. Precisamos arrastar a variável Renda mais vezes até o campo DADOS e modificar a forma de resumir os resultados: basta clicar duas vezes sobre o botão da variável, e escolher a opção, ver Figura 159.

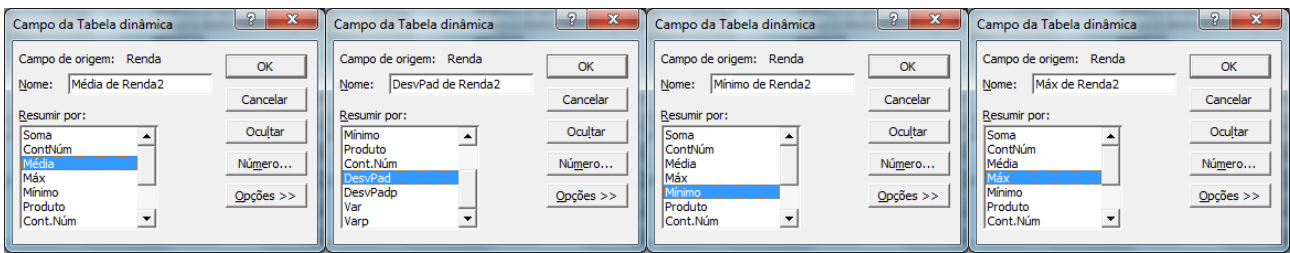


Figura 159 – Formas de resumir os dados de Renda

Usamos “DesvPad” porque os dados do arquivo Celular01.xls referem-se a uma amostra.

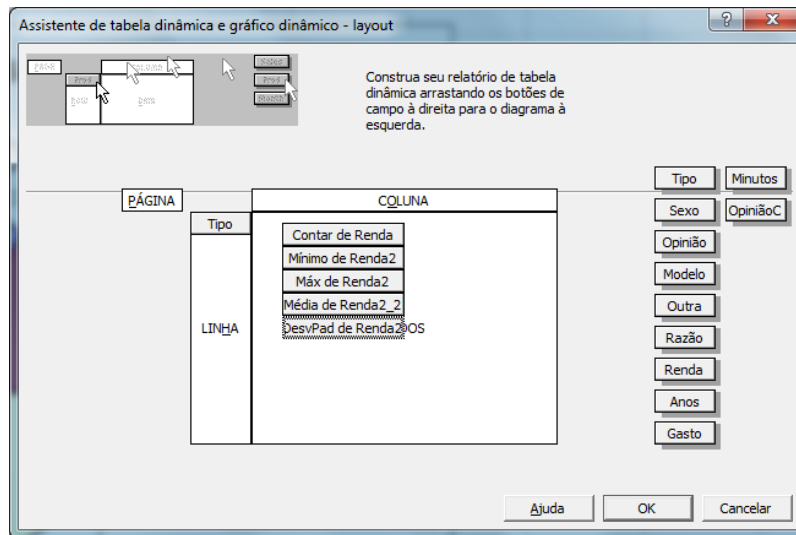


Figura 160 - Layout da tabela dinâmica Renda x Tipo 2a parte

Agora temos o layout pronto. Podemos modificar os nomes na tabela dinâmica e ocultar as células vazias para torná-la mais apresentável, o que resulta na Figura 161.

	BI	BJ	BK
1	Tipo	Dados	Total
2	Pós-pago	Freqüências	2102
3		Mínimo de Renda	1,032
4		Máximo de Renda	43,878
5		Média de Renda	8,383899952
6		Desvio padrão de Renda	5,751495488
7	Pré-pago	Freqüências	2891
8		Mínimo de Renda	1,003
9		Máximo de Renda	35,796
10		Média de Renda	3,918023183
11		Desvio padrão de Renda	3,258436647
12	Total	Freqüências	4993
13	Total	Mínimo de Renda	1,003
14	Total	Máximo de Renda	43,878
15	Total	Média de Renda	5,79693185
16	Total	Desvio padrão de Renda	4,992472753

Figura 161 - Tabela dinâmica de Renda por Tipo

É possível calcular o coeficiente de variação percentual a partir dos dados da tabela, mas a descrição será mais pobre do que a feita na seção 5.3.2, devido à falta de medidas como mediana e quartis (e as derivadas dela).

Tendência central:

- a média de Renda dos clientes de Pós-pago é superior a dos de Pré-pago (mais de 2 vezes) e maior do que a média geral.

Dispersão:

- a variação total da renda é de 1,003 (mínimo) a 43,878 (máximo), sendo que é mais acentuada no Pós-pago (1,032 a 43,878- diferença de 43 vezes) do que no Pré-pago (1,003 a 35,796 – diferença de “apenas” 35 vezes), mas em ambos a dispersão é grande;

- a grande dispersão é confirmada pelo alto valor do coeficiente de variação percentual, indicando que o desvio padrão representa 86,13% da média no total, mas é menor no Pós-pago (68,60%) do que no Pré-pago (83,16%) – os coeficientes são os mesmos mostrados na Figura 155;

Assimetria: não há como avaliar a assimetria.

Valores discrepantes: devido à grande diferença entre as médias e os máximos *presume-se* a

existência de valores discrepantes superiores.

Se quiséssemos acrescentar mais uma variável de agrupamento, Opinião, por exemplo, bastaria arrastá-la para o campo PÁGINA na Figura 160, de forma semelhante ao caso da tabela dinâmica de três variáveis qualitativas (seção 4.3). As medidas seriam apresentadas em função de cada valor de Opinião e do total geral. Isso pode ser muito útil em estudos mais sofisticados. Nada impede a colocação de mais de uma variável no campo PÁGINA, a não ser a crescente dificuldade em interpretar os resultados encontrados. Vamos ver um exemplo, usando a variável Sexo no campo PÁGINA, o layout da tabela seria:

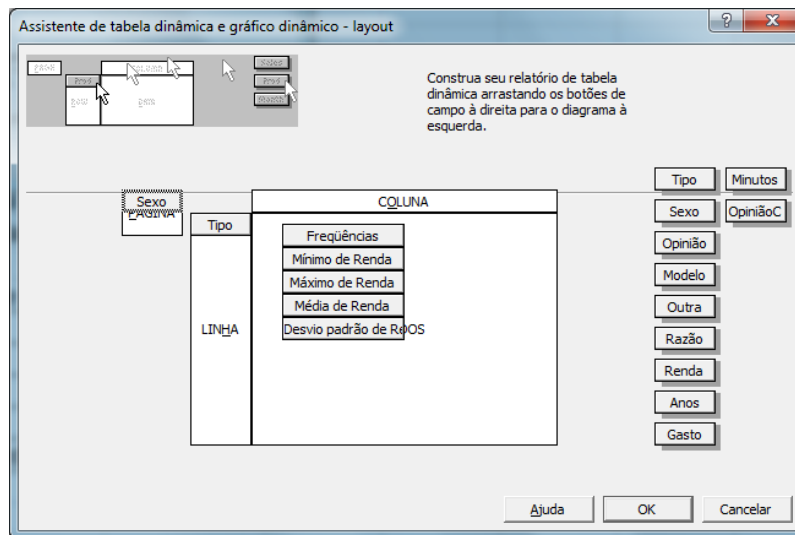


Figura 162 - Layout da tabela dinâmica Renda x Tipo x Sexo

O resultado pode ser visto na Figura 163:

	BI	BJ	BK
1	Sexo	(Tudo)	
2			
3	Tipo	Dados	Total
4	Pós-pago	Freqüências	2102
5		Mínimo de Renda	1,032
6		Máximo de Renda	43,878
7		Média de Renda	8,383899952
8		Desvio padrão de Renda	5,751495488
9	Pré-pago	Freqüências	2891
10		Mínimo de Renda	1,003
11		Máximo de Renda	35,796
12		Média de Renda	3,918023183
13		Desvio padrão de Renda	3,258436647
14	Total Freqüências		4993
15	Total Mínimo de Renda		1,003
16	Total Máximo de Renda		43,878
17	Total Média de Renda		5,79693185
18	Total Desvio padrão de Renda		4,992472753

Figura 163 - Tabela dinâmica de Renda x Tipo x Sexo (geral)

Na Figura 163 a tabela é exatamente igual a da Figura 161 pois não acionamos o filtro da variável Sexo. Na Figura 164 são apresentadas as duas tabelas de Renda por Tipo para o sexo Feminino e Masculino.

	BI	BJ	BK		BI	BJ	BK
1	Sexo	Feminino		1	Sexo	Masculino	
2				2			
3	Tipo	Dados	Total	3	Tipo	Dados	Total
4	Pós-pago	Frequências	709	4	Pós-pago	Frequências	1388
5		Mínimo de Renda	1,032	5		Mínimo de Renda	1,074
6		Máximo de Renda	26,202	6		Máximo de Renda	43,878
7		Média de Renda	6,01730791	7		Média de Renda	9,590888889
8		Desvio padrão de Renda	4,142805999	8		Desvio padrão de Renda	6,078290176
9	Pré-pago	Frequências	1537	9	Pré-pago	Frequências	1352
10		Mínimo de Renda	1,003	10		Mínimo de Renda	1,005
11		Máximo de Renda	14,592	11		Máximo de Renda	35,796
12		Média de Renda	2,906101563	12		Média de Renda	5,070190828
13		Desvio padrão de Renda	2,110273099	13		Desvio padrão de Renda	3,891982437
14	Total Frequências		2246	14	Total Frequências		2740
15	Total Mínimo de Renda		1,003	15	Total Mínimo de Renda		1,005
16	Total Máximo de Renda		26,202	16	Total Máximo de Renda		43,878
17	Total Média de Renda		3,887712121	17	Total Média de Renda		7,358608473
18	Total Desvio padrão de Renda		3,247881088	18	Total Desvio padrão de Renda		5,59309405

Figura 164 - Tabelas dinâmicas de Renda x Tipo x Sexo (Feminino e Masculino)

Podemos calcular os coeficientes de variação percentual e resumir-los em uma tabela:

Geral	86,13%	Pós-pago	68,60%	Feminino geral	83,54%	Masculino geral	76%
		Pré-pago	83,16%	Feminino Pós-pago	68,85%	Masculino Pós-pago	63%
				Feminino Pré-pago	72,62%	Masculino Pré-pago	77%

Tendência central:

- os homens apresentam maior tendência central de Renda, pois sua média vale 7,358 salários mínimos contra 3,8877 das mulheres;
- quando se segmenta por Tipo e Sexo percebe-se que para os dois sexos os clientes de Pós-pago têm maiores médias (6,017 salários mínimos versus 2,906 no feminino, 9,59 versus 3,89 no masculino);
- a subdivisão com maior tendência central de Renda são os homens de Pós-pago, com 9,59 salários mínimos, maior do que a média geral de Pós-pago (8,383 salários mínimos) e do que a média geral de homens (7,358 salários mínimos);
- a subdivisão com menor tendência central de Renda são as mulheres de Pré-pago, com 2,906 salários mínimos, menor do que a média geral de Pré-pago (3,918 salários mínimos) e do que a média geral das mulheres (3,8877 salários mínimos).

Dispersão (através dos coeficientes de variação de Renda, pois as médias são substancialmente diferentes):

- a dispersão do Pré-pago (83,16%) é maior do que a do Pós-pago (68,60%);
- as mulheres têm maior dispersão (83,54%) do que os homens (76%);
- em ambos os sexos a dispersão no Pós-pago é menor do que no Pré-pago, mas nas mulheres de Pós-pago (68,85%) a dispersão é maior do que entre os homens no mesmo tipo (63%), e no Pré-pago a situação se inverte (72,62% contra 77%);

Assimetria: não há como avaliar a assimetria.

Valores discrepantes: novamente, devido à grande diferença entre as médias e os máximos¹⁵ em todas as subdivisões, *presume-se* a existência de valores discrepantes superiores.

¹⁵ A diferença entre as médias e máximos de Renda considerando a subdivisão por Sexo e Tipo vai de 4,35 vezes (Pós-pago Feminino) a 7,06 vezes (Pré-pago Masculino)