

INE 7001 - Procedimentos de Análise Exploratória de Dados utilizando o Excel 2007
Professor Marcelo Menezes Reis

Índice Analítico

1. Filtragem.....	4
1.1 – Filtragem para identificação de células vazias.....	4
1.2 – Filtragem para correção de erros em variáveis qualitativas	7
1.3 - Filtragem para variáveis quantitativas	8
2. Classificação dos Dados	10
3. Recodificação dos dados.....	12
3.1 – Recodificação de Variáveis Qualitativas	12
3.2 – Recodificação de Variáveis Quantitativas	14
4. Análise de dados de variáveis qualitativas – com tabelas dinâmicas.....	19
4.1 – Uma variável qualitativa	19
4.2 – Duas variáveis qualitativas	23
4.3 – Três Variáveis Qualitativas	28
4.4 – Gráficos para variáveis qualitativas	31
4.4.1 – Gráficos para uma variável qualitativa	31
4.4.2 – Gráficos para duas variáveis qualitativas.....	38
5. Análise de dados de variáveis quantitativas.....	41
5.1 - Procedimentos para variáveis quantitativas discretas – tabelas dinâmicas.....	41
5.1.1 – Distribuição de frequência acumulada.....	44
5.1.2 – Histograma para variáveis quantitativas discretas	45
5.2 - Procedimentos para variáveis quantitativas contínuas.....	49
5.2.1 – Histograma para variáveis quantitativas contínuas.....	54
5.3 – Medidas de síntese	55
5.3.1 – Medidas de síntese para uma variável quantitativa	55
5.3.2 – Medidas de síntese para uma variável quantitativa em função de uma qualitativa – por classificação.....	60
5.3.3 – Medidas de síntese para uma variável quantitativa em função de uma qualitativa – por tabela dinâmica	62

Índice de Figuras

Figura 1 - Arquivo Celular01.xlsx	4
Figura 2 - Seleção de variável para filtragem	5
Figura 3 - Variável escolhida para filtragem pelo menu Dados	5
Figura 4 - Variável escolhida para filtragem pelo menu Início	5
Figura 5 - Valores da variável Tipo - obtidos por Filtragem.....	6
Figura 6 - Valores desmarcados Figura 7 - Opção "Vazias" selecionada Figura 8 - Variável Tipo - Vazias	6
Figura 9 - Filtro de texto: personalização	7
Figura 10 Erros de Pós-pago Figura 11 - Células com erros em Pós-pago Figura 12 - Células corrigidas.....	7
Figura 13 - Valores da variável Tipo - sem erros	8
Figura 14 - Filtragem da variável Anos	8
Figura 15 - Filtro de Renda - Menores Valores Figura 16 - Filtro de Renda - Maiores valores.....	9
Figura 17 - Dados sobre o cliente com a maior renda	9
Figura 18 - Filtragem de Renda - dados para 1,005 salários mínimos.....	9
Figura 19 - Classificação pelo menu Dados.....	10
Figura 20 - Classificação pelo menu Início.....	10
Figura 21 - Opções de classificação	10
Figura 22 - Opção de classificação apenas por Tipo	11
Figura 23 - Arquivo de dados classificado em função da variável Tipo (parcial)	11
Figura 24 - Opção de classificação por Tipo e Opinião.....	11
Figura 25 - Arquivo de dados classificado em função das variáveis Tipo e Opinião (parcial).....	12
Figura 26 - Criação da variável OpiniãoC	12
Figura 27 - Função SE para recodificação de Opinião	13
Figura 28 - Função SE para recodificar variável Opinião - considerando células vazias.....	14
Figura 29 - Recodificação da variável Opinião - correspondência das células vazias.....	14
Figura 30 - Função SE para transformar variável Renda - considerando células vazias.....	15
Figura 31 - Transformação da variável Renda	15
Figura 32 - Estrutura de recodificação da variável Renda	16
Figura 33 - Função de recodificação da variável Renda em RendaC	17

Figura 34 - Filtro da variável recodificada RendaC	17
Figura 35 - Variável ClienteAlvo: resultado da recodificação de OpiniãoC e RendaC	18
Figura 36 - Filtro da variável ClienteAlvo.....	18
Figura 37 – Menu de inserção Figura 38 - Seleção de dados de tabela dinâmica	19
Figura 39 - Tela de seleção de intervalo de dados - minimizada.....	19
Figura 40 - Layout da Tabela Dinâmica	20
Figura 41 - Tabela dinâmica de Tipo - apenas frequência	20
Figura 42 - Tabela dinâmica de Tipo.....	21
Figura 43 - Opções do campo Valor.....	21
Figura 44 - Configuração do Campo Valor: Resumir por.....	21
Figura 45 - Configuração do Campo Valor: Mostrar valores como... ..	22
Figura 46 - Configuração do Campo Valor: Mostrar valores como % do total.....	22
Figura 47 - Tabela dinâmica de Tipo: frequências absolutas e relativas – 1ª versão	22
Figura 48 - Tabela dinâmica de Tipo: frequências absolutas e relativas – 2ª versão	23
Figura 49 - Opções de exibição da tabela de Tipo: com remoção das células vazias	23
Figura 50 - Célula de destino de uma tabela dinâmica	23
Figura 51 - Tabela dinâmica de Tipo x Opinião – Layout inicial.....	24
Figura 52 - Tabela dinâmica de Tipo x Opinião – Layout com frequências absolutas e todos os %	24
Figura 53 - Tabela dinâmica de Tipo x Opinião – Layout com frequências absolutas e todos os %, modificado.....	25
Figura 54 - % por linha, coluna e total	25
Figura 55 - Tabela dinâmica de Tipo x Opinião (sem células vazias).....	25
Figura 56 - Tabela dinâmica Tipo x Opinião - opções de formatação.....	26
Figura 57 - Tabela dinâmica Tipo x Opinião - mudança de ordem de valores.....	26
Figura 58 - Tabela dinâmica de Tipo x Opinião.....	26
Figura 59 - Percentuais de Opinião por Tipo de serviço.....	27
Figura 60 - Célula de destino de uma tabela dinâmica	28
Figura 61 - Lista de campos de uma tabela dinâmica de 3 variáveis qualitativas	28
Figura 62 - Tabela dinâmica de Tipo por Opinião em função do Sexo dos clientes.....	29
Figura 63 - Opções de exibição da variável Sexo	29
Figura 64 - Tipo x Opinião para o sexo Feminino.....	29
Figura 65 - Tipo x Opinião para o sexo Masculino.....	29
Figura 66 - Percentuais de Opinião x Sexo Figura 67 - Percentuais de Opinião x Sexo – Pós-pago.....	30
Figura 68 - Percentuais de Opinião x Sexo – Pré-pago Figura 69 - Percentuais de Opinião x Tipo – Feminino.....	30
Figura 70 - Percentuais de Opinião x Tipo –Masculino	31
Figura 71 - Opções de gráficos	31
Figura 72 - Subtipos de gráficos de colunas	32
Figura 73 - Gráfico de colunas de Tipo - 1a versão.....	32
Figura 74 - Campos do gráfico dinâmico Figura 75 - Opções do campo %	33
Figura 76 - Gráfico de colunas de Tipo - 2a versão.....	33
Figura 77 - Opções de design de gráfico dinâmico de colunas	33
Figura 78 - Opções de layout de gráfico dinâmico de colunas.....	33
Figura 79 - Opções de legenda do gráfico dinâmico	34
Figura 80 - Opções de Título do Eixo Vertical Principal em Gráfico Dinâmico.....	34
Figura 81 - Gráfico de colunas de Tipo de serviço - Final.....	35
Figura 82 - Opções de formatação de gráfico e eixo	35
Figura 83 - Opções de Eixo Figura 84 - Mudança de escala no eixo vertical	36
Figura 85 - Gráfico de colunas de Tipo de serviço – escala vertical modificada	36
Figura 86 - Subtipos de gráficos em Pizza.....	37
Figura 87 - Gráfico em pizza de Tipo de serviço - 1a versão.....	37
Figura 88 - Opções de design de gráfico dinâmico em Pizza.....	37
Figura 89 - Opções de layout de gráfico dinâmico em Pizza	37
Figura 90 - Opções de Rótulo de Dados para Gráfico em Pizza	38
Figura 91 - Mais opções de Rótulo de Dados Figura 92 - Gráfico em pizza de Tipo de Serviço - Final.....	38
Figura 93 - Tabela dinâmica de Tipo por Opinião - apenas frequências absolutas	38
Figura 94 - Gráfico de colunas 2D agrupadas de Tipo por Opinião	39
Figura 95 - Gráfico de colunas 2D agrupadas de Tipo por Opinião – Pouco relacionamento.....	39
Figura 96 - Alteração do tipo de gráfico.....	40
Figura 97 - Opções de alteração de tipo de gráfico	40
Figura 98 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião	40
Figura 99 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião – Pouco relacionamento.....	41
Figura 100 - Layout da tabela dinâmica de Anos Figura 101 - Tabela dinâmica de Anos.....	42
Figura 102 - Cópia das células da Tabela dinâmica de Anos.....	42
Figura 103 - Inserção de células no Excel 2007.....	43

Figura 104 - Inserção de células: deslocamento para baixo.....	43
Figura 105 - Células inseridas na Tabela de Anos Figura 106 - Tabela de frequências de Anos.....	43
Figura 107 - Tabela de Anos: frequências e percentuais	44
Figura 108 - CONT.SE: primeira frequência acumulada da variável Ano.....	44
Figura 109 - Tabela de Anos: frequências absolutas, acumuladas e %	45
Figura 110 - Histograma de Anos - 1a versão.....	45
Figura 111 - Opções de gráfico - histograma de Anos	46
Figura 112 - Seleção de fontes de dados - histograma de Anos	46
Figura 113 - Histograma de Anos - 2a versão.....	46
Figura 114 - Seleção de fontes de dados - histograma de Anos - modificações.....	47
Figura 115 - Histograma de Anos - 3a versão.....	47
Figura 116 - Opções de formatação de séries de dados - histograma Anos	47
Figura 117 - Opções de formatar séries de dados.....	48
Figura 118 - Histograma de Anos completos como cliente	48
Figura 119 - Mínimo e máximo para a variável Renda	49
Figura 120 - Intervalo para Renda.....	49
Figura 121 - Amplitude de classes de Renda	49
Figura 122 - Cálculo dos limites inferior e do limite superior da primeira classe de Renda	50
Figura 123 - Cálculo dos limites inferior e superior da segunda classe de Renda	50
Figura 124 - Classes de Renda.....	50
Figura 125 - Ponto médio da primeira classe de Renda.....	51
Figura 126 - Obtenção da frequência da primeira classe de Renda	51
Figura 127 - Obtenção da frequência da última classe de Renda	52
Figura 128 - Obtenção da frequência acumulada da primeira classe de Renda.....	52
Figura 129 - Obtenção da frequência acumulada da primeira classe de Renda.....	53
Figura 130 - Renda agrupada em classes: frequências absolutas, percentuais, frequências acumuladas e percentuais acumulados	53
Figura 131 - Seleção de dados para Histograma de Renda	54
Figura 132 - Histograma da Renda dos clientes em salários mínimos - com limites das classes.....	54
Figura 133 - Fórmulas no Excel 2007	55
Figura 134 - Fórmulas no Excel 2007 - Mais funções	55
Figura 135 - Medidas de síntese de Renda - 1a parte	56
Figura 136 - Funções disponíveis no Excel	56
Figura 137 - Funções estatísticas no Excel - 1a abordagem.....	56
Figura 138 - Argumentos da função Média (para Renda) - 1a versão	57
Figura 139 - Argumentos das funções Mediana e Moda (Renda) - 1a versão.....	57
Figura 140 - Argumentos da função desvio padrão (Renda) - 1a versão	57
Figura 141 - Argumentos das funções Quartil Inferior e Superior (Renda) - 1a versão	58
Figura 142 - Argumentos das funções Mínimo e Máximo (Renda) - 1a versão.....	58
Figura 143 - Medidas de síntese de Renda	58
Figura 144 - Cálculo do CV%.....	58
Figura 145 - Cálculo das diferenças entre os quartis e a mediana	59
Figura 146 - Cálculo dos limites para valores discrepantes	59
Figura 147 - Medidas de síntese para Renda (completas).....	59
Figura 148 - Breakdown de Renda por Tipo - a partir do arquivo classificado.....	61
Figura 149 - Diagrama em caixas de Renda por Tipo	62
Figura 150 - Seleção de dados da tabela dinâmica - Renda por Tipo.....	63
Figura 151 - Layout da tabela dinâmica Renda x Tipo 1a parte.....	63
Figura 152 - Layout da tabela dinâmica Renda x Tipo 2a parte.....	63
Figura 153 - Formas de resumir os dados de Renda	64
Figura 154 - Layout da tabela dinâmica Renda x Tipo 3a parte.....	64
Figura 155 - Tabela dinâmica de Renda por Tipo.....	65
Figura 156 - Layout da tabela dinâmica Renda x Tipo x Sexo	65
Figura 157 - Tabela dinâmica de Renda x Tipo x Sexo (geral).....	66
Figura 158 - Tabelas dinâmicas de Renda x Tipo x Sexo (Feminino e Masculino)	66

INE 7001 - Procedimentos de Análise Exploratória de Dados utilizando o Excel 2007

Professor Marcelo Menezes Reis

O objetivo deste texto é apresentar os principais procedimentos de Análise Exploratória de Dados, tal como apresentados em sala, mas utilizando a planilha eletrônica Excel. Os dados estão no arquivo Celular01.xlsx¹, disponível na páginas das disciplina: contém as informações sobre 5000 clientes da operadora de celular ATCHIM, totalizando 10 variáveis:

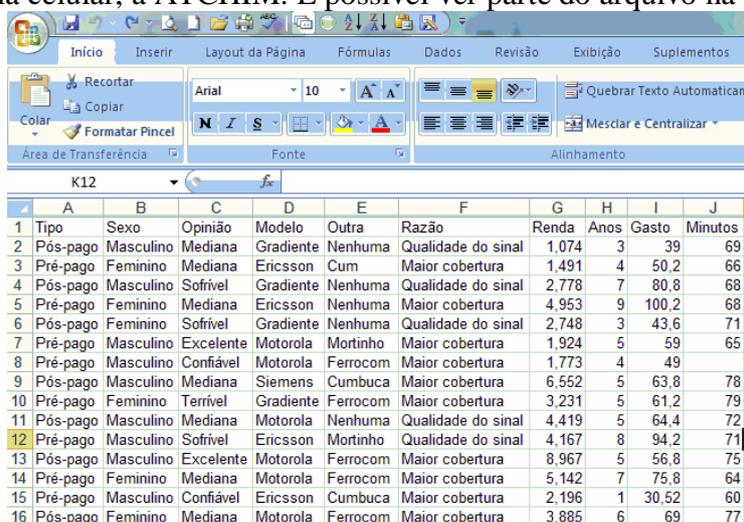
- 1) Tipo de serviço: se o cliente é usuário de celular pré-pago ou pós-pago.
- 2) Sexo do cliente: masculino ou feminino.
- 3) Opinião do cliente sobre a operadora ATCHIM: terrível, sofrível, mediana, confiável, excelente.
- 4) Outra operadora do cliente antes de ser cliente da ATCHIM: Cumbuca, Escuridão, 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.xlsx, 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:



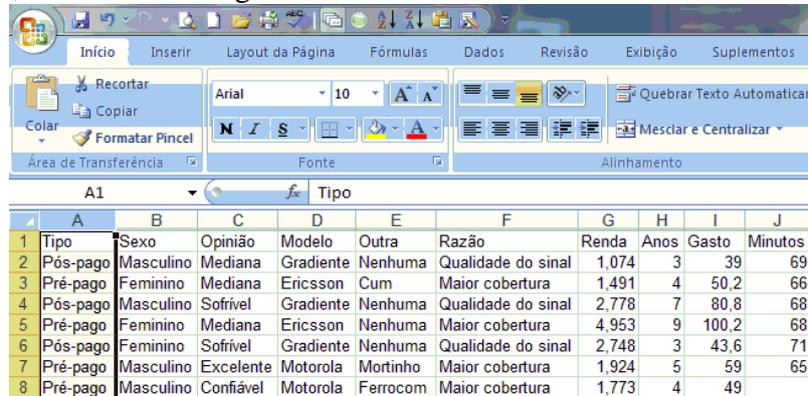
	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69
3	Pré-pago	Feminino	Mediana	Ericsson	Cum	Maior cobertura	1,491	4	50,2	66
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	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	Ferrocom	Maior cobertura	1,773	4	49	
9	Pós-pago	Masculino	Mediana	Siemens	Cumbuca	Maior cobertura	6,552	5	63,8	78
10	Pré-pago	Feminino	Terrível	Gradiente	Ferrocom	Maior cobertura	3,231	5	61,2	79
11	Pós-pago	Masculino	Mediana	Motorola	Nenhuma	Qualidade do sinal	4,419	5	64,4	72
12	Pré-pago	Masculino	Sofrível	Ericsson	Mortinho	Qualidade do sinal	4,167	8	94,2	71
13	Pós-pago	Masculino	Excelente	Motorola	Ferrocom	Maior cobertura	8,967	5	56,8	75
14	Pré-pago	Feminino	Mediana	Motorola	Ferrocom	Maior cobertura	5,142	7	75,8	64
15	Pré-pago	Masculino	Confiável	Ericsson	Cumbuca	Maior cobertura	2,196	1	30,52	60
16	Pós-pago	Feminino	Mediana	Motorola	Ferrocom	Maior cobertura	3,885	6	69	77

Figura 1 - Arquivo Celular01.xlsx

¹ Há também o Celular01.xls, para os procedimentos usando o Excel 2003.

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.

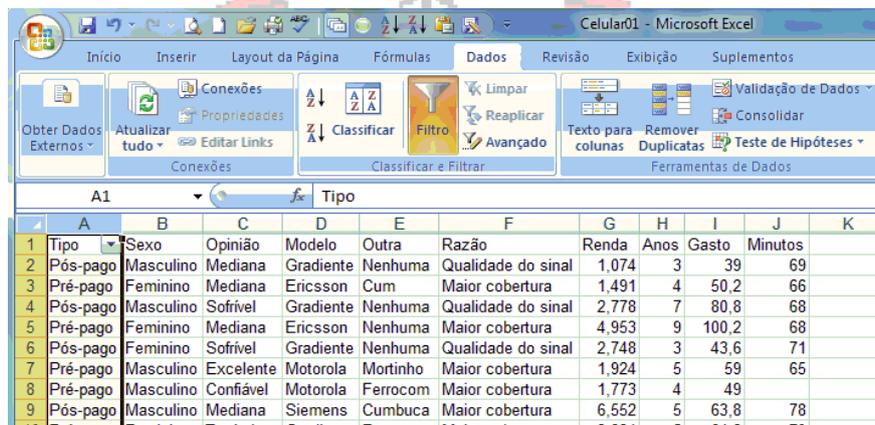


	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2	Pós-pago	Masculino	Mediana	Gradiente	Nenhuma	Qualidade do sinal	1,074	3	39	69
3	Pré-pago	Feminino	Mediana	Ericsson	Cum	Maior cobertura	1,491	4	50,2	66
4	Pós-pago	Masculino	Sofrível	Gradiente	Nenhuma	Qualidade do sinal	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 2 - Seleção de variável para filtragem

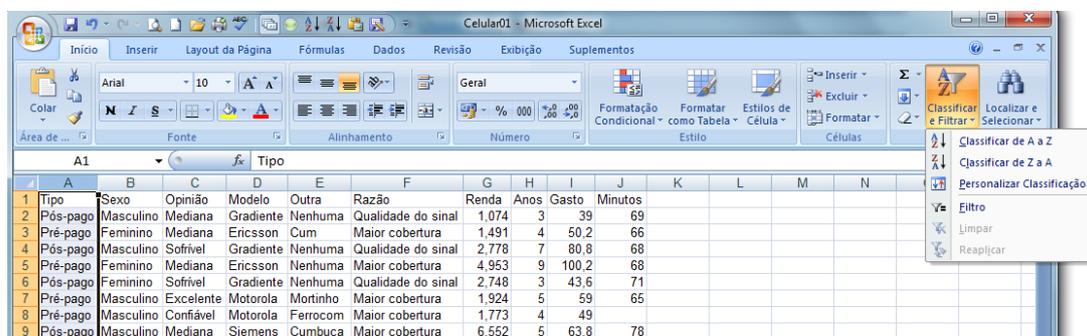
Para realizar a filtragem no Excel 2007 temos dois caminhos:

- tal como no Excel 2003 e anteriores, através do menu Dados, escolher Filtro (Figura 3) – resultando em um botão com uma seta ▼ na coluna de Tipo;
- ou no menu Início, opção Classificar e Filtrar, linha Filtro (Figura 4)², com o mesmo resultado.



	A	B	C	D	E	F	G	H	I	J	K
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	Cum	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		
9	Pós-pago	Masculino	Mediana	Siemens	Cumbuca	Maior cobertura	6,552	5	63,8	78	

Figura 3 - Variável escolhida para filtragem pelo menu Dados



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
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	Cum	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					
9	Pós-pago	Masculino	Mediana	Siemens	Cumbuca	Maior cobertura	6,552	5	63,8	78				

Figura 4 - Variável escolhida para filtragem pelo menu Início

² Aliás, se o cursor for posicionado em qualquer célula das colunas A a J todas as variáveis terão o botão de filtro adicionado.

Independente do caminho a variável estará com Auto-Filtro implementado. No Auto-Filtro o Excel automaticamente seleciona todos os valores possíveis para a variável, incluindo células vazias e erros. Se pressionarmos o botão com a seta ao lado do nome da variável vamos obter a Figura 5:

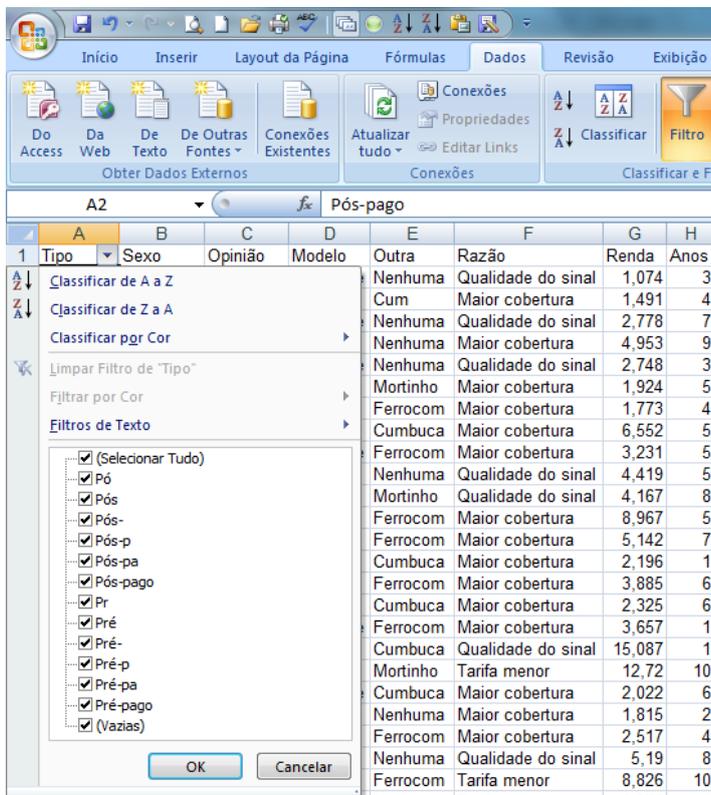


Figura 5 - Valores da variável Tipo - obtidos por Filtragem

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.

Se quisermos exibir apenas as células “Vazias” primeiramente devemos desmarcar a opção “Selecionar Tudo”, vista na Figura 5, resultando na Figura 6.

Para mostrar apenas as células vazias, basta marcar “Vazias”, na Figura 7, obtendo o resultado na Figura 8.

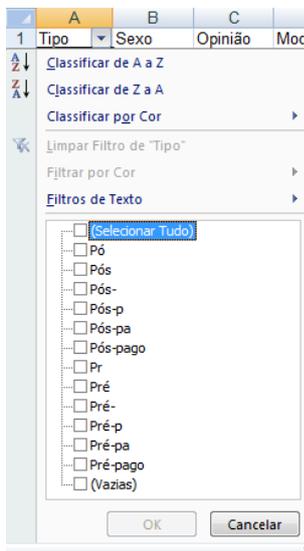


Figura 6 -Valores desmarcados

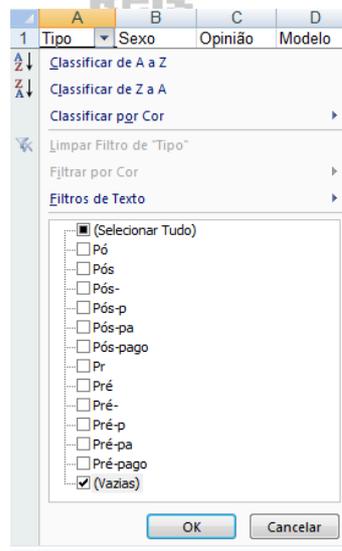


Figura 7 - Opção "Vazias" selecionada

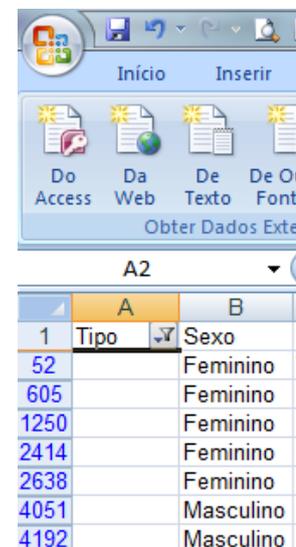


Figura 8 - Variável Tipo - Vazias

Os dados perdidos encontram-se nas células A52, A605, A1250, A2414, A2638, A4051 e A4192, totalizando 7 observações. Como são apenas 7 em 5000 registros, representando 0,14%, bem abaixo de 5%, podemos considerar uma quantidade aceitável. Porém, se possível, devemos investigar as causas deste “sumiço de dados”, e se possível, descobrir seu conteúdo (deve haver algum registro

do cliente 52 em algum lugar, que diga qual o seu tipo de serviço). Este procedimento precisa ser feito para todas as variáveis (qualitativas e quantitativas).

Uma outra opção de filtragem possível é usar um “Filtro de texto” personalizado: fornecemos ao Excel as características procuradas para selecionar as células a exibir, selecionando “Filtro de Texto” na Figura 5 obtemos a Figura 9:

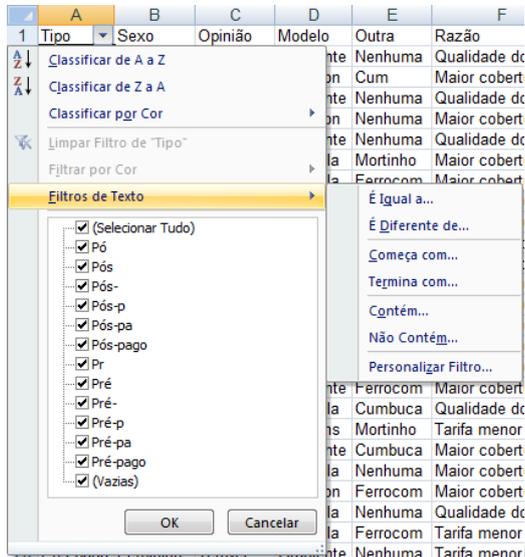


Figura 9 - Filtro de texto: personalização

Há várias possibilidades de personalização, permitindo escolher exibir as células que atendem a vários critérios: é igual a, é diferente de, começa com, termina com, contém, não contém, e ainda uma opção de “Personalizar Filtro” – quer permite usar como referência células de qualquer planilha do Excel.

Outro uso da Filtragem é a identificação de erros de registro (ver Figura 5), e a sua correção.

1.2 – Filtragem para correção de erros em variáveis qualitativas

Para corrigir os valores incorretamente registrados de Pós-pago devemos desmarcar a opção “Selecione tudo” (ver Figura 6) e então selecionar todas as incorretas, como na Figura 10. O Excel apresentará apenas as células que contém os valores incorretos (Figura 11). São 17 células apenas. Não obstante, precisamos corrigi-las. Basta digitar o valor correto na primeira célula do intervalo (A19 na Figura 11). Em seguida, “arrasta-se” o valor correto até a última célula com o valor incorreto, e todos serão corrigidos, como mostrado na Figura 12.



Figura 10 Erros de Pós-pago

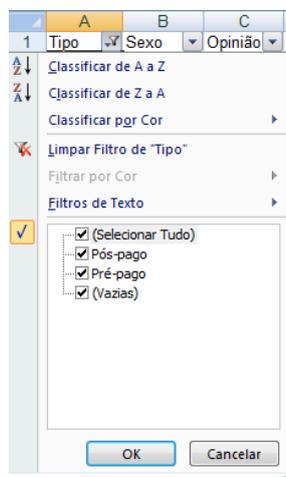
	A	B	C	D
1	Tipo	Sexo	Opinião	Modelo
19	Pós-pa	Masculino	Confiável	Motorola
234	Pó	Feminino	Terrível	Ericsson
426	Pó	Feminino	Terrível	Motorola
439	Pó	Masculino	Terrível	Siemens
573	Pós-	Masculino	Terrível	Gradiente
698	Pós-pa	Masculino	Terrível	Siemens
988	Pós	Masculino	Sofrível	Samsung
1018	Pós	Feminino	Sofrível	Motorola
1337	Pós	Masculino	Mediana	Samsung
1961	Pós-p	Masculino	Mediana	Samsung
2416	Pós-pa	Feminino	Sofrível	Ericsson
3276	Pós-pa	Feminino	Sofrível	Motorola
3415	Pós-pa	Masculino	Sofrível	Siemens
3662	Pós-	Masculino	Terrível	Siemens
3903	Pó	Feminino	Sofrível	Motorola
4024	Pós-pa	Masculino	Sofrível	Siemens
4579	Pós-p	Masculino	Mediana	Gradiente

Figura 11 - Células com erros em Pós-pago

	A	B	C
1	Tipo	Sexo	Opinião
19	Pós-pago	Masculino	Confiável
234	Pós-pago	Feminino	Terrível
426	Pós-pago	Feminino	Terrível
439	Pós-pago	Masculino	Terrível
573	Pós-pago	Masculino	Terrível
698	Pós-pago	Masculino	Terrível
988	Pós-pago	Masculino	Sofrível
1018	Pós-pago	Feminino	Sofrível
1337	Pós-pago	Masculino	Mediana
1961	Pós-pago	Masculino	Mediana
2416	Pós-pago	Feminino	Sofrível
3276	Pós-pago	Feminino	Sofrível
3415	Pós-pago	Masculino	Sofrível
3662	Pós-pago	Masculino	Terrível
3903	Pós-pago	Feminino	Sofrível
4024	Pós-pago	Masculino	Sofrível
4579	Pós-pago	Masculino	Mediana

Figura 12 - Células corrigidas

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 13:



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

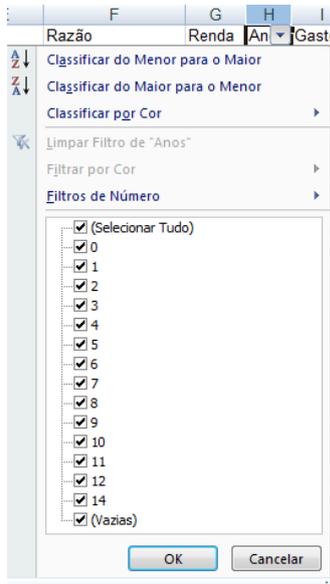
Observe que com um procedimento simples é possível corrigir todos os erros de registro, avaliar a quantidade de dados perdidos e eventualmente (através da opção Filtro de Texto) realizar outras filtragens.

Figura 13 - Valores da variável Tipo - sem erros

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 Filtro (como para a variável Tipo, seja pelo menu Dados – Figura 3, ou Início – Figura 4), chegamos à tela mostrada na Figura 14.



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.

A filtragem é especialmente importante no caso de variáveis quantitativas contínuas. 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 15, e o máximo (43,878 salários mínimos), mostrado na Figura 16.

Figura 14 - Filtragem da variável Anos

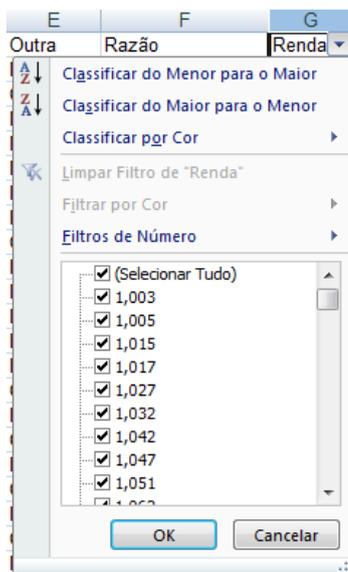


Figura 15 - Filtro de Renda - Menores Valores

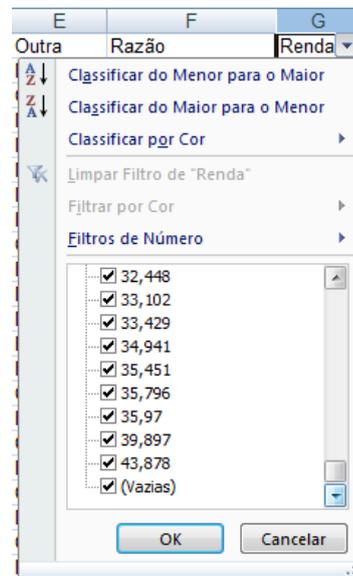


Figura 16 - Filtro de Renda - Maiores valores

Se quisermos saber mais sobre o cliente com a maior renda basta desmarcar “Selecionar Tudo” e selecionar o valor 43,878 e teremos a tela da Figura 17, com todas as informações disponíveis sobre este cliente.

	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos
2242	Pós-pago	Masculino		Samsung	Nenhuma	Tarifa menor	43,878	6	68,4	109

Figura 17 - Dados sobre o cliente com a maior renda

As informações estão armazenadas na linha 2242 da planilha. 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 em média por mês. 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 18.

	A	B	C	D	E	F	G	H	I	J
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	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 18 - Filtragem de Renda - dados para 1,005 salários mínimos

As informações sobre os clientes que percebem 1,005 salários mínimos estão armazenadas nas linhas 793, 2359, 2418, 2836, 2924, 4137 e 4773. Eles 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.

A filtragem para variáveis quantitativas é muito útil para identificar os extremos do conjunto, o que permitirá a construção de distribuições de frequências, o que será visto posteriormente.

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. É preciso que o cursor esteja em alguma das células dos dados.

Para realizar a filtragem no Excel 2007 temos dois caminhos:

- tal como no Excel 2003 e anteriores, através do menu Dados, escolher Classificar (Figura 19) – resultando na Figura 21;
- ou no menu Início, opção Classificar e Filtrar, linha “Personalizar Classificação” (Figura 20), com o mesmo resultado.

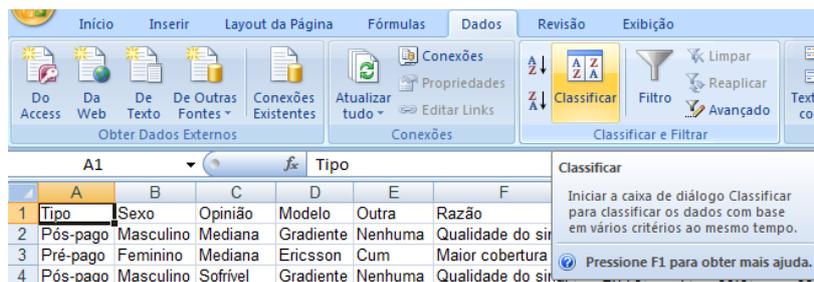


Figura 19 - Classificação pelo menu Dados

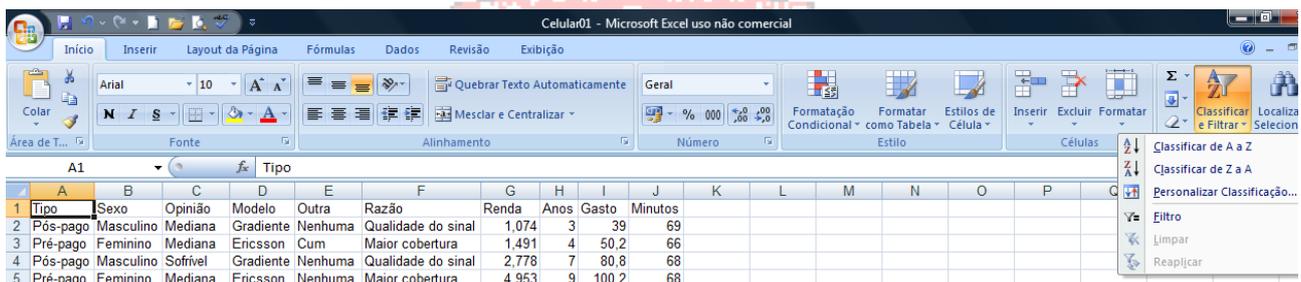


Figura 20 - Classificação pelo menu Início

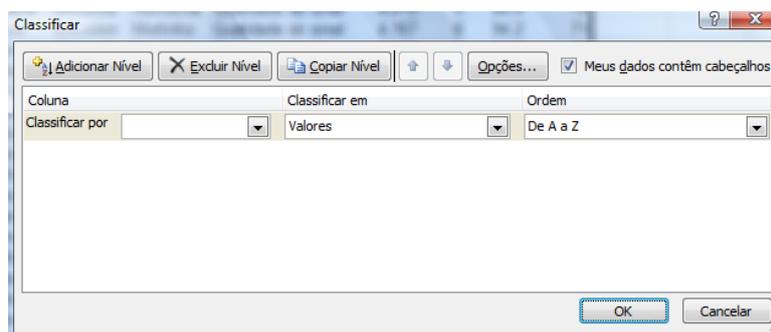


Figura 21 - Opções de classificação

O padrão do Excel 2007 é a classificação por apenas uma variável. Mas, é permitido adicionar mais variáveis através do botão “Adicionar Nível”, o que torna possível ordenar os dados em função dos valores de várias variáveis (no Excel 2003 há a possibilidade de usar apenas 3 variáveis). Se desejássemos fazer a classificação apenas em função de Tipo bastaria escolhê-la na primeira opção da tela da Figura 22.

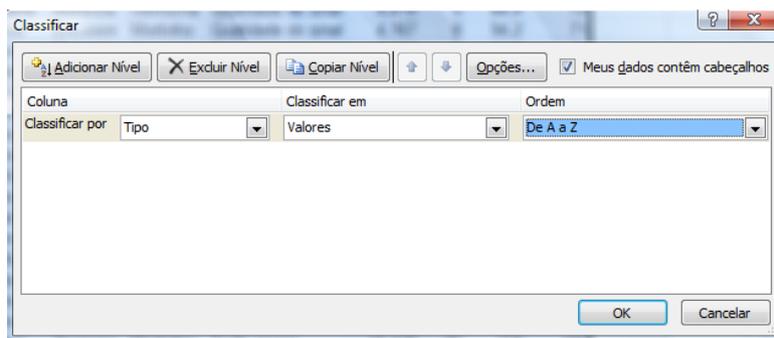


Figura 22 - Opção de classificação apenas por Tipo

É possível fazer a classificação pelos valores de Tipo (ou outro aspecto, como a cor da célula, por exemplo), e em ordem de A a Z – ascendente (ou de Z-A – descendente, ou de forma personalizada). Vamos manter a classificação pelos valores, de A a Z, o que resulta na Figura 23.

	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 23 - 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 20) 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 23), 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 19 ou 20 e depois a 21, com a adição de um nível, mas agora escolhendo Opinião na segunda opção de classificação, tal como mostrado na Figura 24.

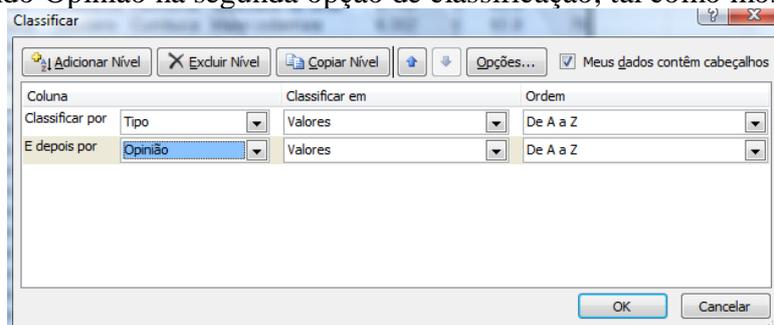
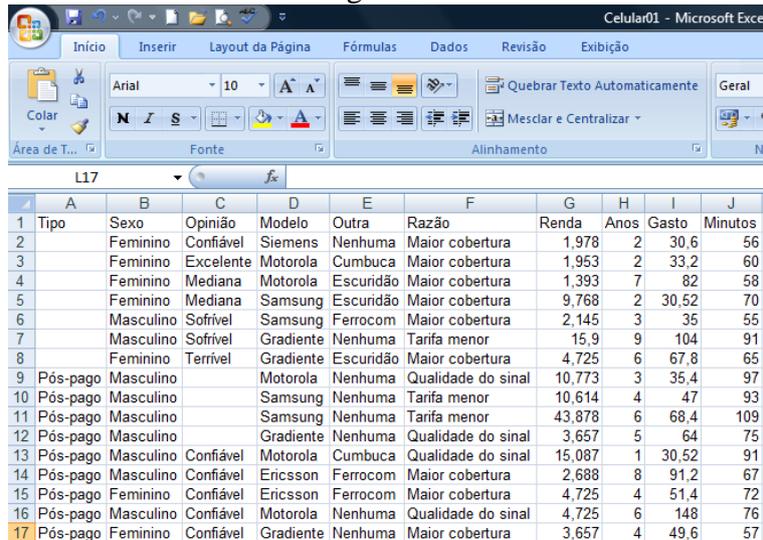


Figura 24 - Opção de classificação por Tipo e Opinião

O arquivo classificado é mostrado na Figura 25.



	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	Ferocom	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	Ferocom	Maior cobertura	2,688	8	91,2	67
15	Pós-pago	Feminino	Confiável	Ericsson	Ferocom	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

Figura 25 - 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.

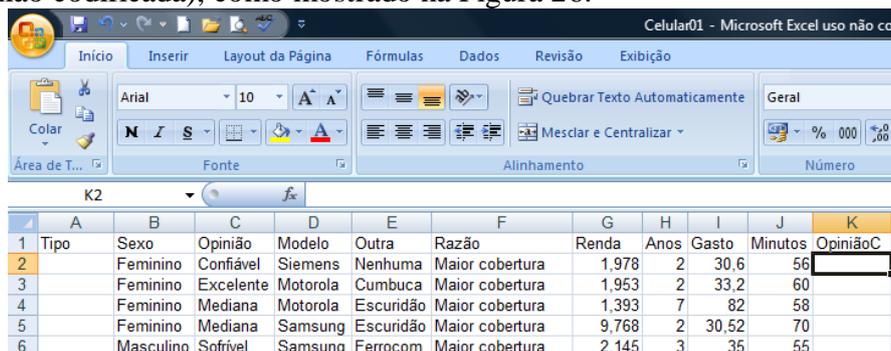
3. Recodificação dos dados

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 selecionar a célula K1, onde será colocado o nome da nova variável, OpiniãoC (opinião codificada), como mostrado na Figura 26.



	A	B	C	D	E	F	G	H	I	J	K
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC
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	Ferocom	Maior cobertura	2,145	3	35	55	

Figura 26 - 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 27 a seguir ilustra o procedimento que queremos fazer.



Figura 27 - 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.

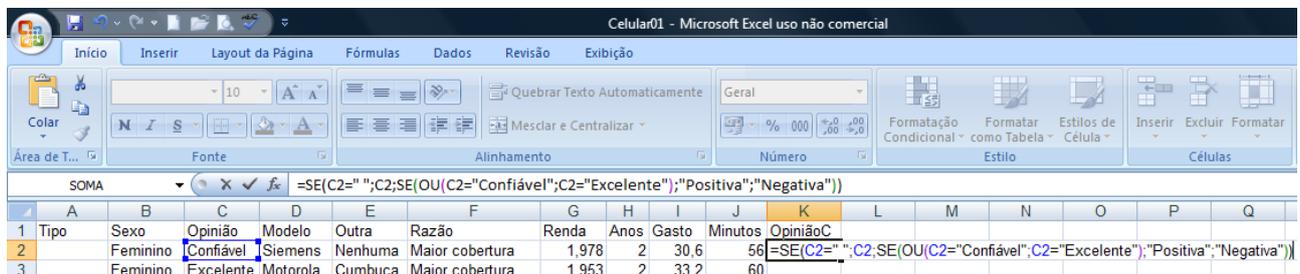
No caso da recodificação da variável Opinião vamos usar uma função SE e dentro dela, no teste lógico uma função OU: SE Opinião for Confiável OU Excelente, OpiniãoC será Positiva, SENÃO será Negativa. Mas, como há células vazias na variável Opinião, se realizarmos a recodificação tal como mostrada aqui “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 28.

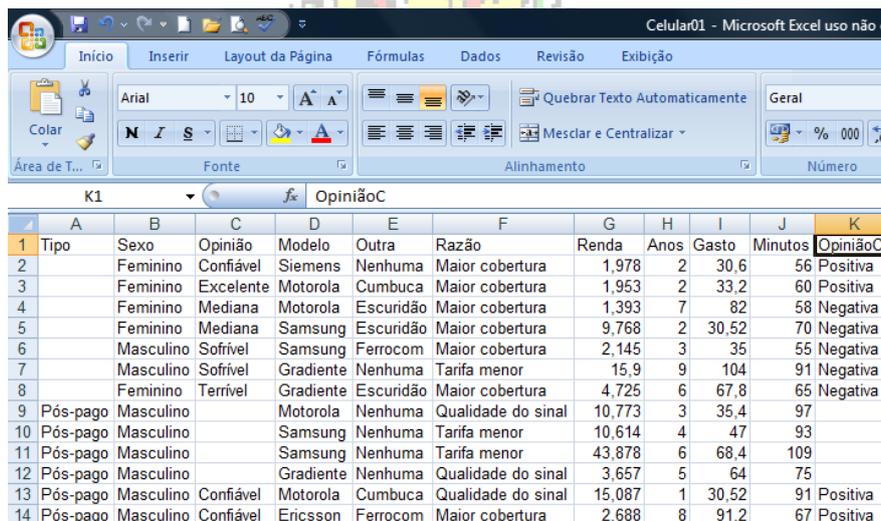


The screenshot shows the Excel interface with the formula bar containing: `=SE(C2="" ;C2;SE(OU(C2="Confiável";C2="Excelente");"Positiva";"Negativa"))`. The spreadsheet below shows the following data:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC						
2		Feminino	Confiável	Siemens	Nenhuma	Maior cobertura	1,978	2	30,6	56	=SE(C2="" ;C2;SE(OU(C2="Confiável";C2="Excelente");"Positiva";"Negativa"))						
3		Feminino	Excelente	Motorola	Cumbuca	Maior cobertura	1,953	2	33,2	60							

Figura 28 - 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 29.



The screenshot shows the Excel interface with the formula bar containing: `OpiniãoC`. The spreadsheet below shows the filtered data:

	A	B	C	D	E	F	G	H	I	J	K
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC
2		Feminino	Confiável	Siemens	Nenhuma	Maior cobertura	1,978	2	30,6	56	Positiva
3		Feminino	Excelente	Motorola	Cumbuca	Maior cobertura	1,953	2	33,2	60	Positiva
4		Feminino	Mediana	Motorola	Escuridão	Maior cobertura	1,393	7	82	58	Negativa
5		Feminino	Mediana	Samsung	Escuridão	Maior cobertura	9,768	2	30,52	70	Negativa
6		Masculino	Sofrível	Samsung	Ferocom	Maior cobertura	2,145	3	35	55	Negativa
7		Masculino	Sofrível	Gradiente	Nenhuma	Tarifa menor	15,9	9	104	91	Negativa
8		Feminino	Terrível	Gradiente	Escuridão	Maior cobertura	4,725	6	67,8	65	Negativa
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	Positiva
14	Pós-pago	Masculino	Confiável	Ericsson	Ferocom	Maior cobertura	2,688	8	91,2	67	Positiva

Figura 29 - 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.

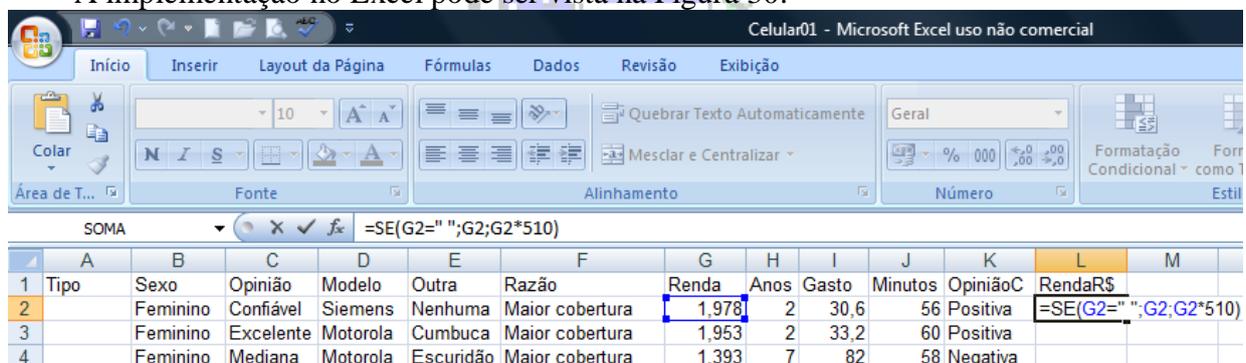
Tais ações facilitarão o estudo da Renda, posto que algumas pessoas têm dificuldade de raciocinar em termos de salários mínimos, e outras preferem uma visão “categorizada” das quantidades.

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)

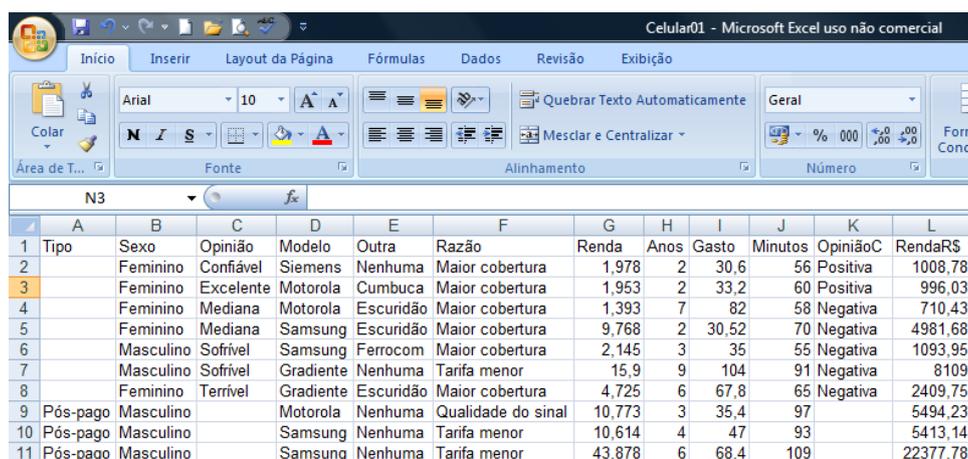
A implementação no Excel pode ser vista na Figura 30.



	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		Feminino	Confiável	Siemens	Nenhuma	Maior cobertura	1,978	2	30,6	56	Positiva	=SE(G2="" ;G2;G2*510)	
3		Feminino	Excelente	Motorola	Cumbuca	Maior cobertura	1,953	2	33,2	60	Positiva		
4		Feminino	Mediana	Motorola	Escuridão	Maior cobertura	1,393	7	82	58	Negativa		

Figura 30 - 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 31.



	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\$
2		Feminino	Confiável	Siemens	Nenhuma	Maior cobertura	1,978	2	30,6	56	Positiva	1008,78
3		Feminino	Excelente	Motorola	Cumbuca	Maior cobertura	1,953	2	33,2	60	Positiva	996,03
4		Feminino	Mediana	Motorola	Escuridão	Maior cobertura	1,393	7	82	58	Negativa	710,43
5		Feminino	Mediana	Samsung	Escuridão	Maior cobertura	9,768	2	30,52	70	Negativa	4981,68
6		Masculino	Sofrível	Samsung	Ferocom	Maior cobertura	2,145	3	35	55	Negativa	1093,95
7		Masculino	Sofrível	Gradiente	Nenhuma	Tarifa menor	15,9	9	104	91	Negativa	8109
8		Feminino	Terrível	Gradiente	Escuridão	Maior cobertura	4,725	6	67,8	65	Negativa	2409,75
9	Pós-pago	Masculino		Motorola	Nenhuma	Qualidade do sinal	10,773	3	35,4	97		5494,23
10	Pós-pago	Masculino		Samsung	Nenhuma	Tarifa menor	10,614	4	47	93		5413,14
11	Pós-pago	Masculino		Samsung	Nenhuma	Tarifa menor	43,878	6	68,4	109		22377,78

Figura 31 - Transformação da variável Renda

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 32.

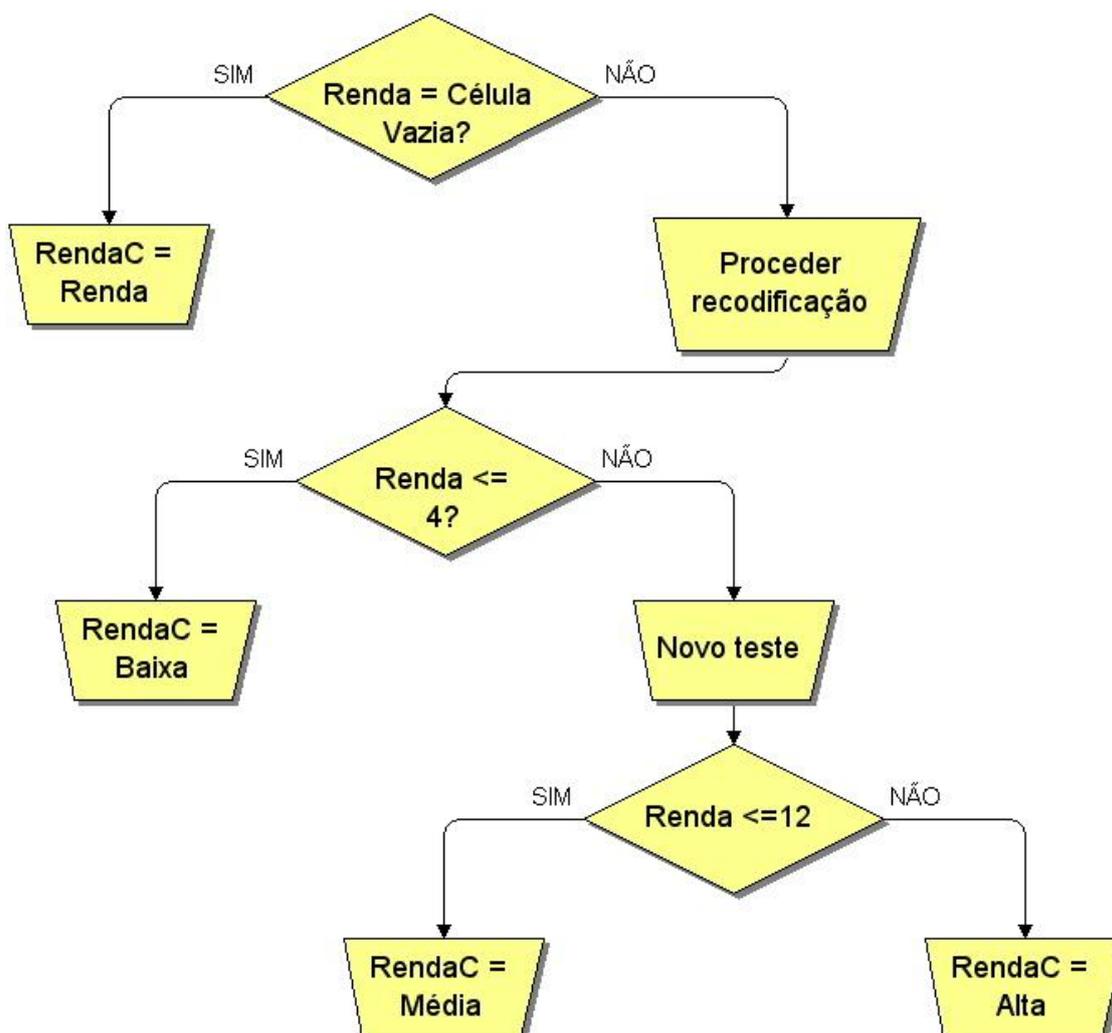


Figura 32 - Estrutura de recodificação da variável Renda

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 33).

	G	H	I	J	K	L	M	N	O	P	Q	R
1	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	RendaC					
2	1,978	2	30,6	56	Positiva	1008,78	=SE(G2="" ;G2;SE(G2<=4;"Baixa";SE(G2<=12;"Média";"Alta"))					
3	1,953	2	33,2	60	Positiva	996,03						

Figura 33 - 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 34.

	G	H	I	J	K	L	M
1	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	RendaC
2	1,978	2	30,6				
3	1,953	2	33,2				
4	1,393	7	82				
5	9,768	2	30,52				
6	2,145	3	35				
7	15,9	9	104				
8	4,725	6	67,8				
9	10,773	3	35,4				
10	10,614	4	47				
11	43,878	6	68,4				
12	3,657	5	64				
13	15,087	1	30,52				
14	2,688	8	91,2				
15	4,725	4	51,4				
16	4,725	6	148				
17	3,657	4	49,6				
18	5,688	6	70,4				
19	1,134	4	45				
20	1,452	7	78,8				
21	2,325	7	78,4				
22	13,173	3	37,4				

Figura 34 - 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 35.

	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	RendaC	ClienteAlvo						
2	1,978	2	30,6	56	Positiva	1008,78	Baixa	Sim						
3	1,953	2	33,2	60	Positiva	996,03	Baixa							
4	1,393	7	82	58	Negativa	710,43	Baixa							
5	9,768	2	30,52	70	Negativa	4981,68	Média							
6	2,145	3	35	55	Negativa	1093,95	Baixa							

Figura 35 - 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 36.

	G	H	I	J	K	L	M	N
1	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	RendaC	ClienteAlvo
2	1,978	2	30,6	56	Positiva			
3	1,953	2	33,2	60	Positiva			
4	1,393	7	82	58	Negativa			
5	9,768	2	30,52	70	Negativa			
6	2,145	3	35	55	Negativa			
7	15,9	9	104	91	Negativa			
8	4,725	6	67,8	65	Negativa			
9	10,773	3	35,4	97				
10	10,614	4	47	93				
11	43,878	6	68,4	109				
12	3,657	5	64	75				
13	15,087	1	30,52	91	Positiva			
14	2,688	8	91,2	67	Positiva			
15	4,725	4	51,4	72	Positiva			
16	4,725	6	148	76	Positiva			
17	3,657	4	49,6	57	Positiva			
18	5,688	6	70,4	66	Positiva			
19	1,134	4	45	66	Positiva			
20	1,452	7	78,8	53	Positiva			
21	2,325	7	78,4	69	Positiva			
22	13,173	3	37,4	75	Positiva			

Figura 36 - 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

Independente se as variáveis de interesse são qualitativas ou quantitativas o procedimento para construção da tabela dinâmica é o mesmo. Posicione o cursor em algum ponto do conjunto de dados, e procure no menu Inserir por Tabela Dinâmica, como na Figura 37.

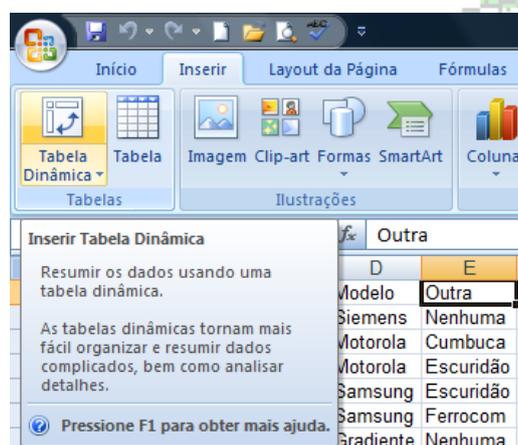


Figura 37 – Menu de inserção

Pressionando Tabela Dinâmica o Excel 2007 apresenta o resultado da Figura 38.

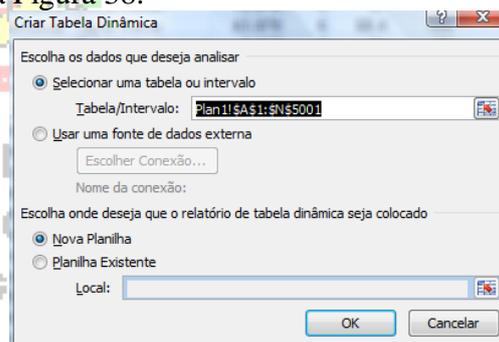


Figura 38 - Seleção de dados de tabela dinâmica

Ao pôr o cursor em alguma parte do banco de dados o Excel 2007 seleciona automaticamente todas as variáveis adjacentes: no presente caso todas as células compreendidas entre A1 e N5001: da variável Tipo até a ClienteAlvo. Pode-se escolher, eventualmente, uma fonte externa ou dados de outra planilha ou outro intervalo de dados na mesma planilha. Para fazer isso basta pressionar o botão com seta vermelha na linha “Tabela/Intervalo” da Figura 38, resultando na Figura 39:

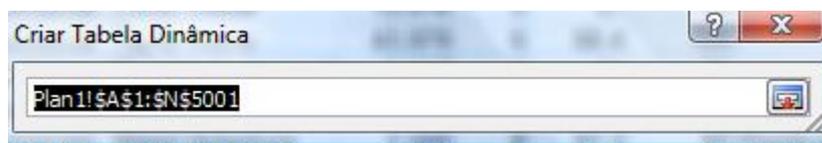


Figura 39 - Tela de seleção de intervalo de dados - minimizada

Com a seleção neste formato você pode procurar o intervalo desejado e selecioná-lo utilizando o mouse. Para retornar à situação da Figura 38 basta pressionar a pequena seta vermelha novamente. Por conveniência, vamos pôr a tabela dinâmica em uma nova planilha. Então, pressionando OK, chega-se à Figura 40.

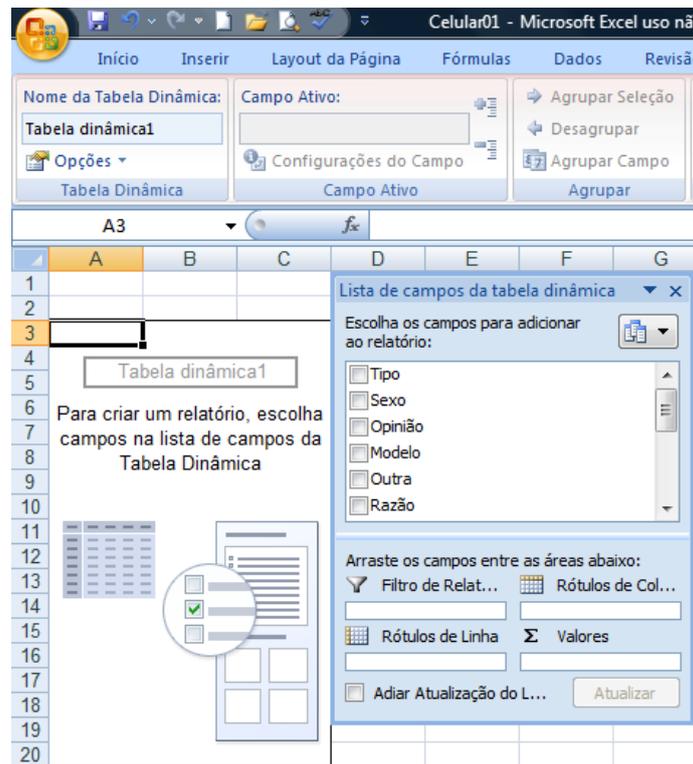


Figura 40 - Layout da Tabela Dinâmica

A versão 2007 do Excel automaticamente cria a tabela, vazia, com os campos disponíveis para adição à direita. Observe à direita os nomes das variáveis existentes no arquivo de dados, e que foram selecionadas na etapa anterior. Vamos imaginar que há interesse apenas na variável Tipo, devemos selecioná-la e arrastá-la até o campo "Rótulos de Linha", ou "Rótulos de Coluna". As outras variáveis não farão parte da tabela. Devemos também arrastar a variável Tipo para o campo "Valores". O resultado pode ser visto na Figura 41.

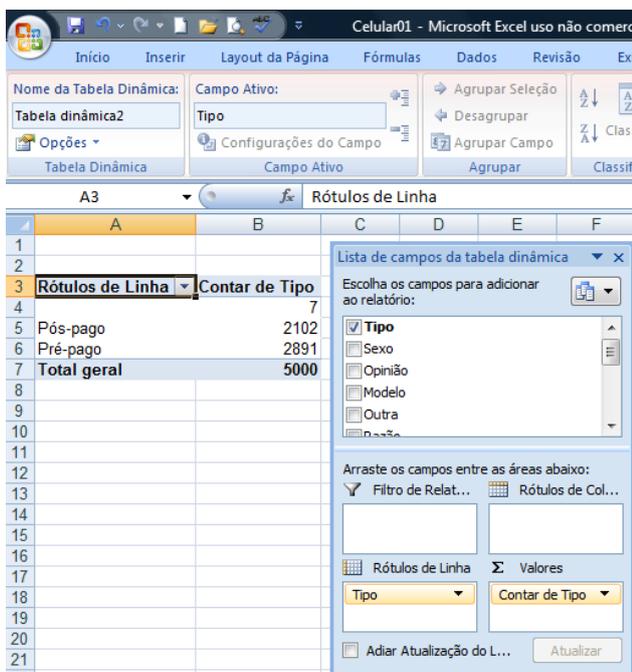


Figura 41 - Tabela dinâmica de Tipo - apenas frequência

Observe que a tabela contém apenas as frequências absolutas de cada valor, além das células vazias. Mas, usualmente é imprescindível apresentar as frequências relativas também, para que seja possível fazer comparações com outros conjuntos de dados. Para tanto devemos arrastar novamente Tipo para o campo Valores, o que nos levará à Figura 42: as frequências absolutas dos valores serão apresentadas mais uma vez, e precisamos modificá-las para que a tabela contenha as frequências absolutas e as relativas. Para fazer isso, devemos pressionar a seta ▼ do segundo botão do campo Valores, o que leva às opções da Figura 43.

Rótulos de Linha	Contar de Tipo	Contar de Tipo2
	7	7
Pós-pago	2102	2102
Pré-pago	2891	2891
Total geral	5000	5000

Figura 42 - Tabela dinâmica de Tipo

Rótulos de Linha	Contar de Tipo	Contar de Tipo2
	7	7
Pós-pago	2102	2102
Pré-pago	2891	2891
Total geral	5000	5000

Figura 43 - Opções do campo Valor

Selecionando “Configurações do Campo de Valor...” é possível mudar o formato de apresentação dos dados, com as várias opções mostradas na Figura 44.

Configurações do Campo de Valor

Nome da Fonte: Tipo

Nome Personalizado: Contar de Tipo2

Resumir por: Mostrar valores como

Resumir campo de valor por

Escolha o tipo de cálculo que deseja usar para resumir os dados do campo selecionado

- Soma
- ContNúm**
- Média
- Máx
- Mínimo
- Produto

Formato do Número

OK Cancelar

Pode-se mudar o tipo de cálculo que se deseja usar para resumir os dados do campo selecionado. Quando a variável é qualitativa a opção padrão é “ContNúm”, contagem. Mas, se a variável for quantitativa, pode-se escolher soma, média, Máx (máximo), Mínimo, Produto e outras funções não visíveis na Figura 44 (desvio padrão amostral e populacional, variância amostral e populacional). É possível ver também a aba “Mostrar valores como”, que uma vez selecionada resulta na Figura 45.

Figura 44 - Configuração do Campo Valor: Resumir por...

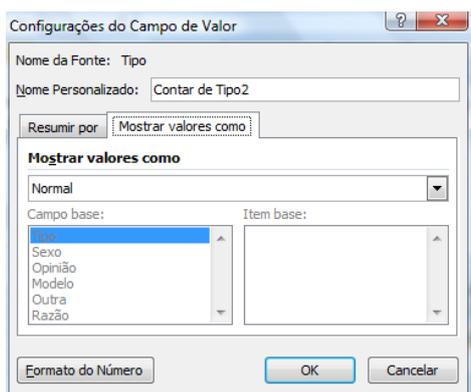


Figura 45 - Configuração do Campo Valor: Mostrar valores como...

A opção padrão é apresentar os valores como “Normal”, ou seja, no caso de contagem para variáveis qualitativas, as frequências absolutas. Mas, pressionando a seta podemos observar as várias opções disponíveis na Figura 46: % de diferença de, soma acumulada, % da linha, % da coluna, % do total entre outras. Como se deseja apresentar as frequências relativas dos valores da variável Tipo, podemos escolher % do total e pressionar OK, a tabela resultante pode ser analisada na Figura 47.

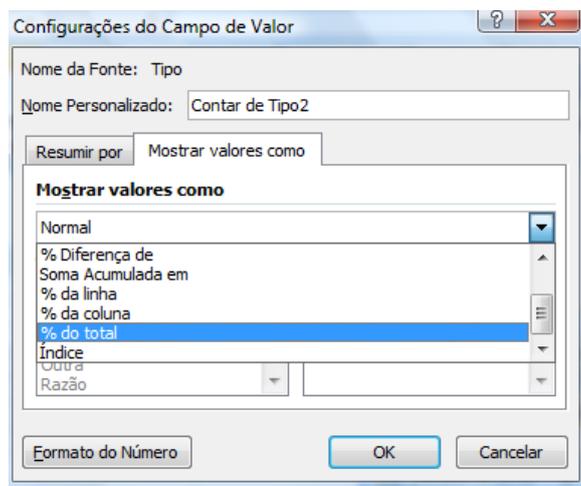


Figura 46 - Configuração do Campo Valor: Mostrar valores como % do total

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

Figura 47 - Tabela dinâmica de Tipo: frequências absolutas e relativas – 1ª versão

As células vazias representam 0,14% do total, perfazendo apenas 7 clientes: é um percentual pequeno (aceita-se até 5%), mas que merece uma reflexão, pois é um dado imprescindível, a empresa precisa descobrir o Tipo de serviço destes clientes, até para fins de tarifação. Percebe-se a preponderância dos clientes do Tipo pré-pago, que representam 57,82% do total (2891 clientes). Observa-se também que a tabela apresenta algumas “poluições”, o que pode confundir sua interpretação por leigos: há uma linha para células vazias, “Rótulos de Linha”, “Valores”, “Contar de Tipo”, “Contar de Tipo2”. Todos estes podem ser modificados, os últimos simplesmente por digitação dos nomes de nosso interesse, resultando na Figura 48.

	A	B	C
1			
2	Frequências		
3	Tipo	Absolutas	%
4		7	0,14%
5	Pós-pago	2102	42,04%
6	Pré-pago	2891	57,82%
7	Total geral	5000	100,00%

Figura 48 - Tabela dinâmica de Tipo: frequências absolutas e relativas – 2ª versão

A remoção da linha com as células vazias, porque sua influência é muito reduzida, pode ser feita pressionando a seta ▼ ao lado de Tipo, levando às opções da Figura 49.

	A	B	C
1			
2	Frequências		
3	Tipo	Absolutas	%
4	Pós-pago	2102	42,10%
5	Pré-pago	2891	57,90%
6	Total geral	4993	100,00%

Figura 49 - Opções de exibição da tabela de Tipo: com remoção das células vazias

Basta desmarcar a opção de exibição das células vazias, pressionar OK, e obtém-se a tabela final, com os valores percentuais já recalculados em relação ao novo total (4993).

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 anterior: posicione o cursor em algum ponto do conjunto de dados, e procure no menu Inserir por Tabela Dinâmica, como na Figura 37. Mas, ao invés de pôr a tabela em uma nova planilha vamos escolher uma célula da mesma planilha onde está a tabela de Tipo, como mostrado na Figura 50.

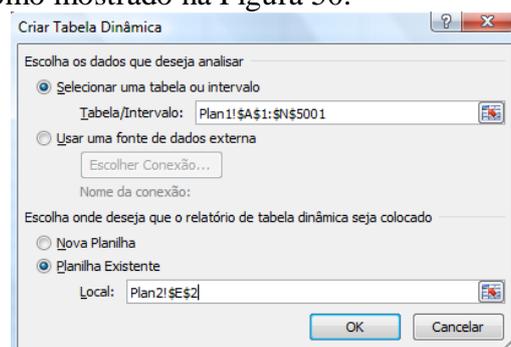


Figura 50 - Célula de destino de uma tabela dinâmica

Pressionando OK, chega-se à uma situação parecida a da Figura 40. Mas, agora vamos colocar duas variáveis na Tabela, Tipo na Linha, e Opinião na Coluna, e apenas uma delas (qualquer uma) em Valores. Como na Figura 51.

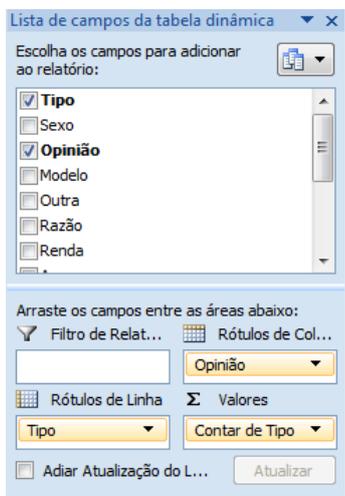


Figura 51 - Tabela dinâmica de Tipo x Opinião – Layout inicial

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 “Valores”, Figura 52. Posteriormente, podemos modificar as configurações do campo Valor, com as devidas adaptações, tal como nas Figuras 43 a 46. Ver Figura 54. A tabela final resultante está na Figura 55, já com as modificações apropriadas (incluindo a remoção das células vazias das duas variáveis).

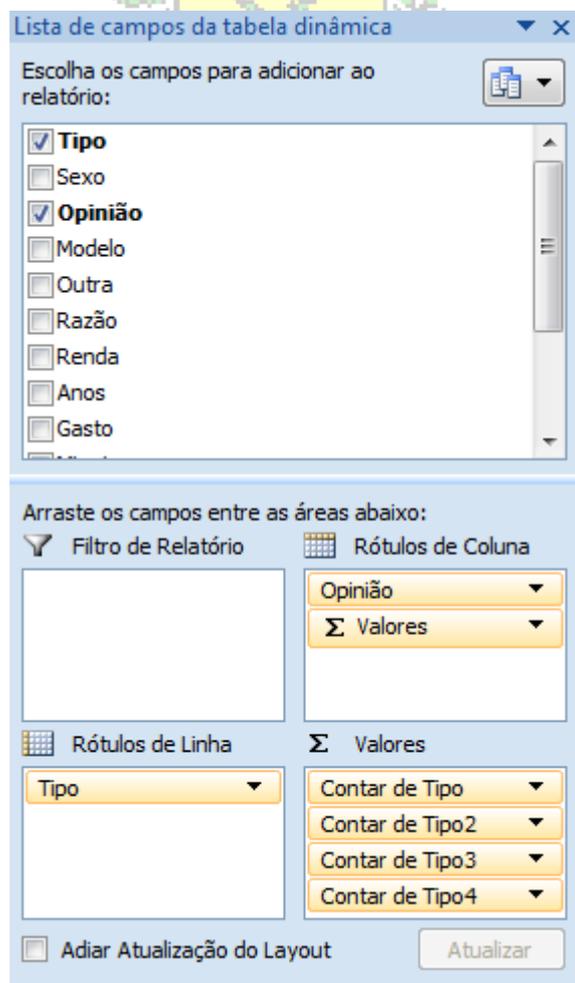


Figura 52 - Tabela dinâmica de Tipo x Opinião – Layout com frequências absolutas e todos os %.

Da forma como está o layout da Figura 52 a tabela não ficará com um bom aspecto, dificultando a sua interpretação. Precisamos mover o botão Σ Valores do campo Rótulos de Coluna (ver Figura 52) para o campo Rótulos de Linha, resultando na Figura 53.

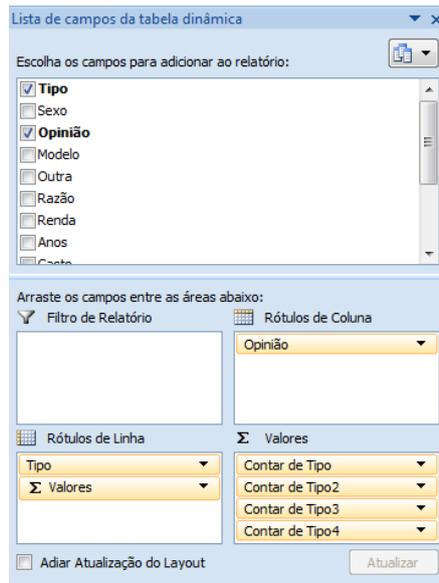


Figura 53 - Tabela dinâmica de Tipo x Opinião – Layout com frequências absolutas e todos os %, modificado

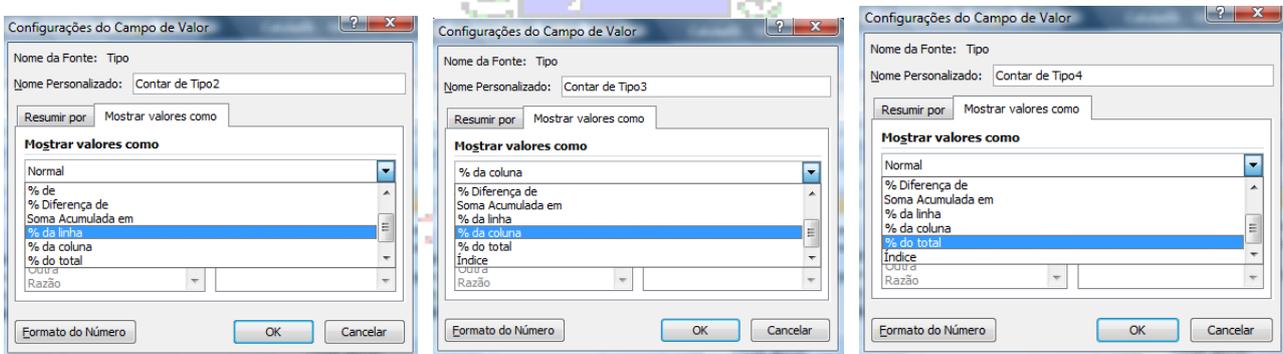


Figura 54 - % por linha, coluna e total

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

Figura 55 - 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 56.

Tipo	Opinião	Excelente	Mediana	Sofrível	Terrível	Total geral
Pós-pago						
Frequência				519	842	2098
% por linha	4,9%			24,74%	40,13%	100,00%
% por coluna	13,4%			51,85%	67,31%	42,10%
% por total	2,0%			10,42%	16,90%	42,10%
Pré-pago						
Frequência				482	409	2885
% por linha	23,0%			16,71%	14,18%	100,00%
% por coluna	86,5%			48,15%	32,69%	57,90%
% por total	13,3%			9,67%	8,21%	57,90%
Total Frequência				1001	1251	4983
Total % por linha	15,3%			20,09%	25,11%	100,00%
Total % por coluna	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Total % por total	15,39%	10,13%	29,28%	20,09%	25,11%	100,00%

Figura 56 - Tabela dinâmica Tipo x Opinião - opções de formatação

Observe a opção Mover na Figura 56, selecionando-a vamos à Figura 57.

Tipo	Opinião	Excelente	Mediana	Sofrível	Terrível	Total geral
Pós-pago						
Frequência				519	842	2098
% por linha	4,9%			24,74%	40,13%	100,00%
% por coluna	13,4%			51,85%	67,31%	42,10%
% por total	2,0%			10,42%	16,90%	42,10%
Pré-pago						
Frequência				482	409	2885
% por linha	23,0%			16,71%	14,18%	100,00%
% por coluna	86,5%			48,15%	32,69%	57,90%
% por total	13,3%			9,67%	8,21%	57,90%
Total Frequência				1001	1251	4983
Total % por linha	15,3%			20,09%	25,11%	100,00%
Total % por coluna	100,00%	100,00%	100,00%	100,00%	100,00%	100,00%
Total % por total	15,39%	10,13%	29,28%	20,09%	25,11%	100,00%

Figura 57 - Tabela dinâmica Tipo x Opinião - mudança de ordem de valores

Há três opções possíveis: mover Opinião para Linhas (o que não queremos), mover “Confíavel” para a direita ou mover para o fim. Como precisamos mudar apenas Confíavel escolhemos mover para a direita, veja a Figura 58.

Tipo	Opinião	Excelente	Confíavel	Mediana	Sofrível	Terrível	Total geral
Pós-pago							
Frequência		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%
% por total		2,75%	2,07%	9,97%	10,42%	16,90%	42,10%
Pré-pago							
Frequência		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%
% por total		7,39%	13,33%	19,31%	9,67%	8,21%	57,90%
Total Frequência		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 % por total		10,13%	15,39%	29,28%	20,09%	25,11%	100,00%

Figura 58 - Tabela dinâmica de Tipo x Opinião

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 Total³);
- 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%⁴), mostrando claramente que estes últimos estão mais insatisfeitos do que aqueles, corroborando as conclusões da análise dos percentuais por linhas. A Figura 59 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.

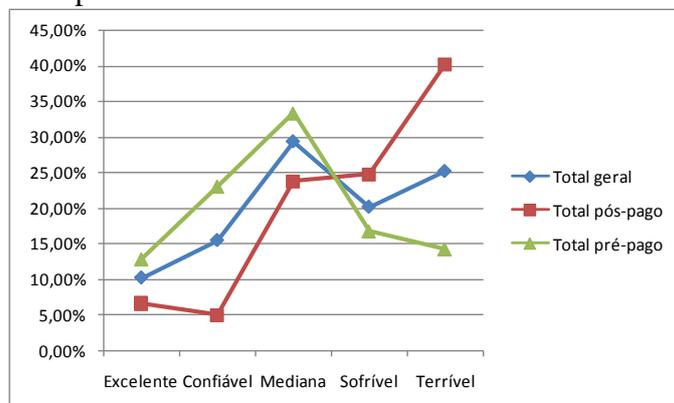


Figura 59⁵ - Percentuais de Opinião por Tipo de serviço

³ Se não houvesse relação entre Tipo e Opinião os percentuais de Pós-pago e Pré-pago deveriam ser próximos de 15,39%.

⁴ 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.

⁵ Esta figura é um gráfico de linhas, que estudaremos posteriormente.

4.3 – Três Variáveis Qualitativas

Em tese poderíamos construir tabelas para cruzar qualquer número de variáveis qualitativas. Contudo, a interpretação dos resultados seria bastante difícil⁶. Mas, para os casos em que há no máximo 3 variáveis é plenamente possível construir e interpretar uma tabela dinâmica no Excel 2007. 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 2007 seguindo os passos usuais de montagem de tabela dinâmica, conforme visto na seção 4.2, para duas variáveis qualitativas:

- posicione o cursor em algum ponto do conjunto de dados; procure no menu Inserir por Tabela Dinâmica, como na Figura 37; ponha a tabela em uma célula da mesma planilha onde está a tabela de Tipo, e a tabela de Tipo por Opinião, como mostrado na Figura 60.

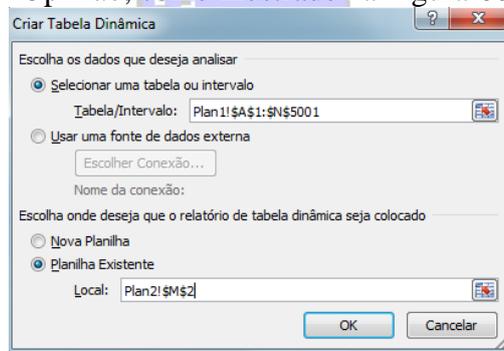


Figura 60 - Célula de destino de uma tabela dinâmica

Pressionando OK, chega-se à uma situação parecida a da Figura 40. Mas, agora vamos colocar *três* variáveis na Tabela, Tipo na Linha, Opinião na Coluna, tal como na seção 4.3, e Sexo no campo Filtro de Relatório: com isso a tabulação cruzada de Tipo e Opinião será construída em função dos valores de Sexo. Não podemos esquecer também do campo Valores: vamos pôr as frequências absolutas e os percentuais por linha⁷ (veja a seção 4.3), resultando na Figura 61.

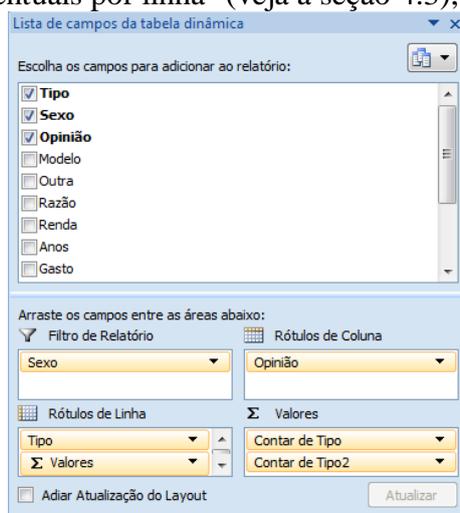


Figura 61 - Lista de campos de uma tabela dinâmica de 3 variáveis qualitativas

⁶ Nestes casos é recomendável usar análise estatística multivariada.

⁷ Vamos manter apenas estes para não sobrecarregar demais a tabela, mas nada impede de manter os percentuais por coluna e em relação ao total geral.

Note-se que o botão Σ Valores já foi movido do campo Rótulos de Coluna (ver Figura 52) para o campo Rótulos de Linha. Após fazer as modificações necessárias (mudança de Contar de Tipo para Freqüências, Contar de Tipo2 para % por linha, mudança de ordem de Confiável, remoção das células vazias de Tipo e de Opinião) o resultado é a Figura 62.

Sexo		(Tudo)					
Rótulos de Coluna		Excelente					
Rótulos de Linha	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 62 - Tabela dinâmica de Tipo por Opinião em função do Sexo dos clientes

Trata-se de parte da tabela mostrada na Figura 58, porque na variável Sexo está selecionada a opção Tudo. Clicando sobre ▼ ao lado de Sexo, vemos as opções de exibição da Figura 63.

Sexo		(Tudo)					
Rótulos de Linha		Excelente					
Rótulos de Linha	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 63 - Opções de exibição da variável Sexo

As possibilidades são: células vazias (-), Feminino e Masculino. Se selecionarmos Feminino, o resultado será a Figura 64. Ao escolhermos Masculino obtemos a Figura 65.

Sexo		Feminino					
Rótulos de Coluna		Excelente					
Rótulos de Linha	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 64 - Tipo x Opinião para o sexo Feminino

Sexo		Masculino					
Rótulos de Coluna		Excelente					
Rótulos de Linha	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 65 - Tipo x Opinião para o sexo Masculino

Repare que na Figura 64 o total geral vale 2241 (e não mais 4983 da Figura 62), pois são consideradas apenas as clientes. Na Figura 65, 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.

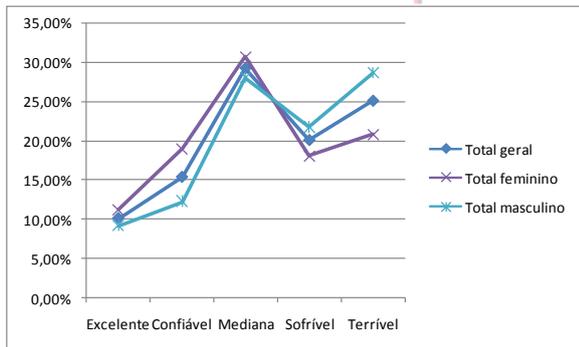


Figura 66 - Percentuais de Opinião x Sexo

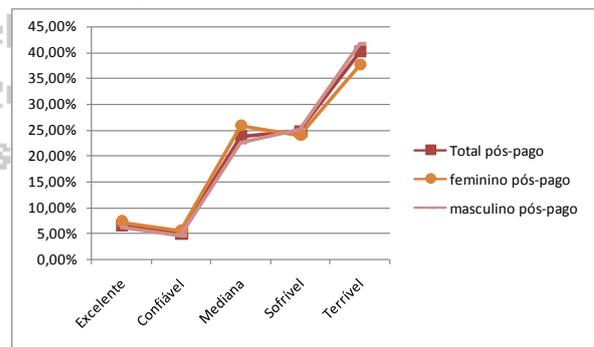


Figura 67 - Percentuais de Opinião x Sexo – Pós-pago

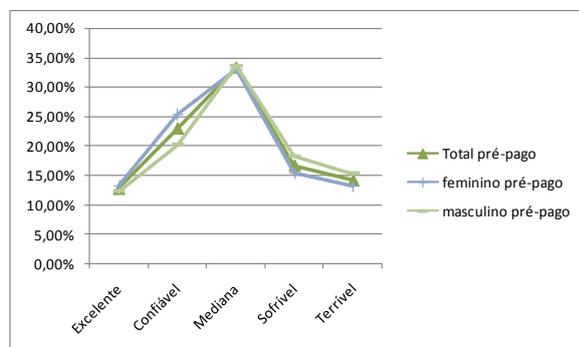


Figura 68 - Percentuais de Opinião x Sexo – Pré-pago

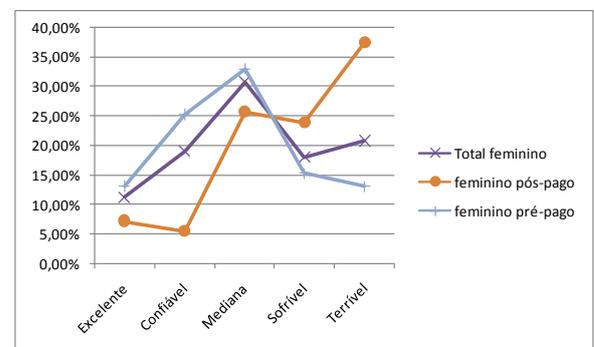


Figura 69 - Percentuais de Opinião x Tipo – Feminino

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 66, 67 e 68). 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 59, 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 69 e 70).

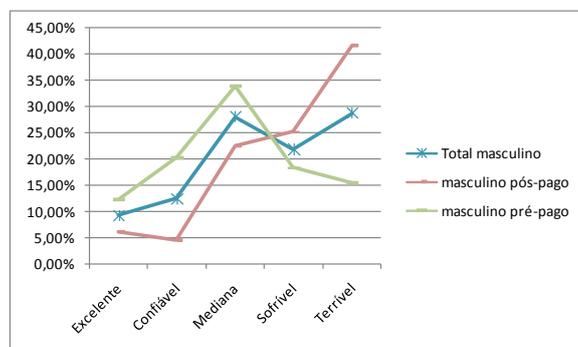


Figura 70 - Percentuais de Opinião x Tipo –Masculino

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 frequências 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.

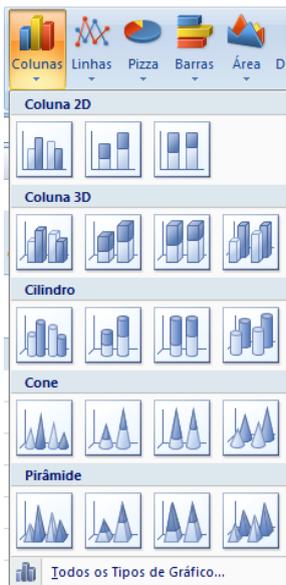
Gráfico de barras (colunas)

Ponha o cursor em algum ponto da tabela dinâmica de Tipo, e procure no menu Inserir por gráficos, como na Figura 71.



Figura 71 - Opções de gráficos

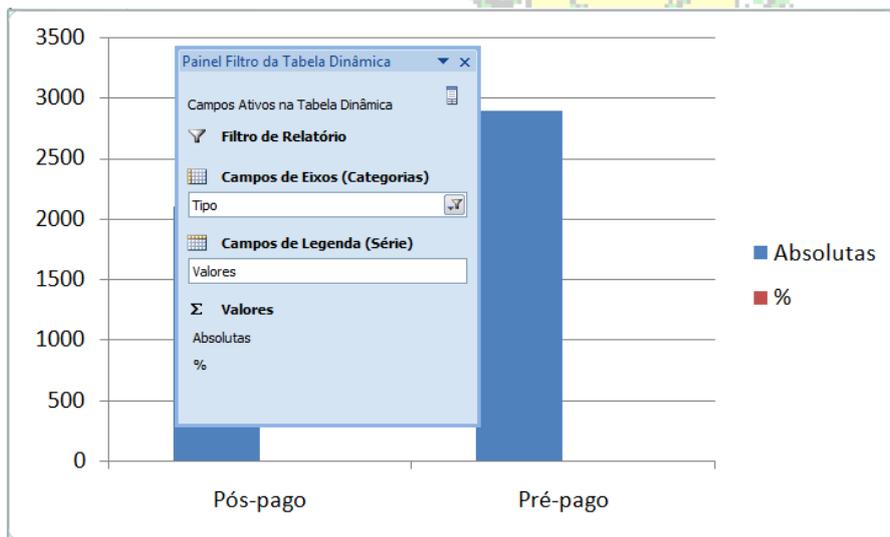
Clique no tipo de gráfico e, em seguida, clique no subtipo de gráfico que deseja usar. No nosso caso, queremos “Colunas” (barras verticais). Se pressionarmos a seta imediatamente abaixo de “Colunas” é possível ver todos os subtipos disponíveis (Figura 72).



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

Para a tabela dinâmica de Tipo vamos escolher Colunas 2D, colunas agrupadas, resultando na Figura 73.

Figura 72 - Subtipos de gráficos de colunas



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

Figura 73 - Gráfico de colunas de Tipo - 1ª versão

Na parte superior de “Painel Filtro da Tabela Dinâmica” há “Campos Ativos na Tabela Dinâmica”: precisamos pressioná-lo para remover os % do gráfico, com a finalidade de torná-lo mais claro. **Mas, ao fazer isso vamos removê-los da tabela dinâmica original também.** Assim, se não quiser perdê-la, copie a tabela e cole-a *como valores* em algum outro ponto da planilha, e construa o gráfico a partir da cópia. Pressionando “Campos Ativos na Tabela Dinâmica”, chegamos à Figura 74:

- selecionando % no campo Σ Valores, vamos obter a figura 75;
- escolhendo “Remover campo”, desaparecerá % do gráfico e da tabela dinâmica original de Tipo, resultando na Figura 76.

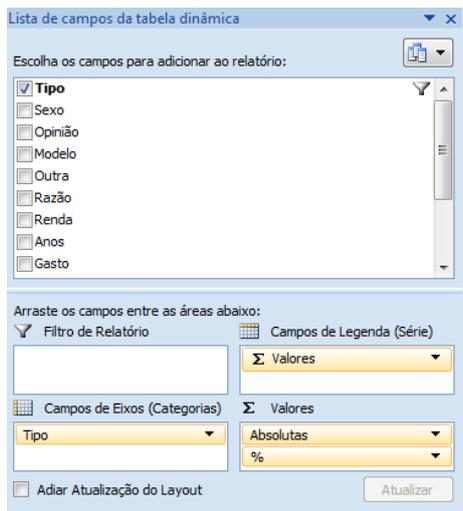


Figura 74 - Campos do gráfico dinâmico

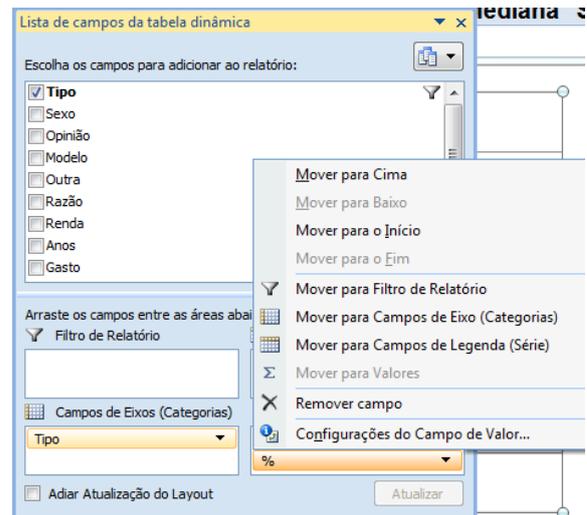


Figura 75 - Opções do campo %

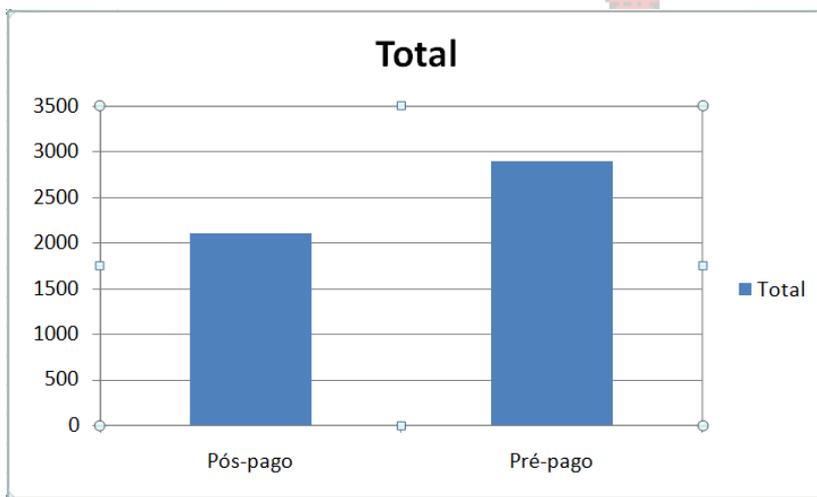


Figura 76 - Gráfico de colunas de Tipo - 2a versão

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

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

As opções mais interessantes neste momento são “Design” (Figura 77) e “Layout” (Figura 78).



Figura 77 - Opções de design de gráfico dinâmico de colunas

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



Figura 78 - Opções de layout de gráfico dinâmico de colunas

Nas opções de layout podemos modificar formato e posição dos títulos (do gráfico e dos eixos), a área de plotagem, adicionar Linha de Tendência e modificar a Legenda. No nosso caso, não queremos exibi-la, então, ao selecionar Legenda obtemos a Figura 79.

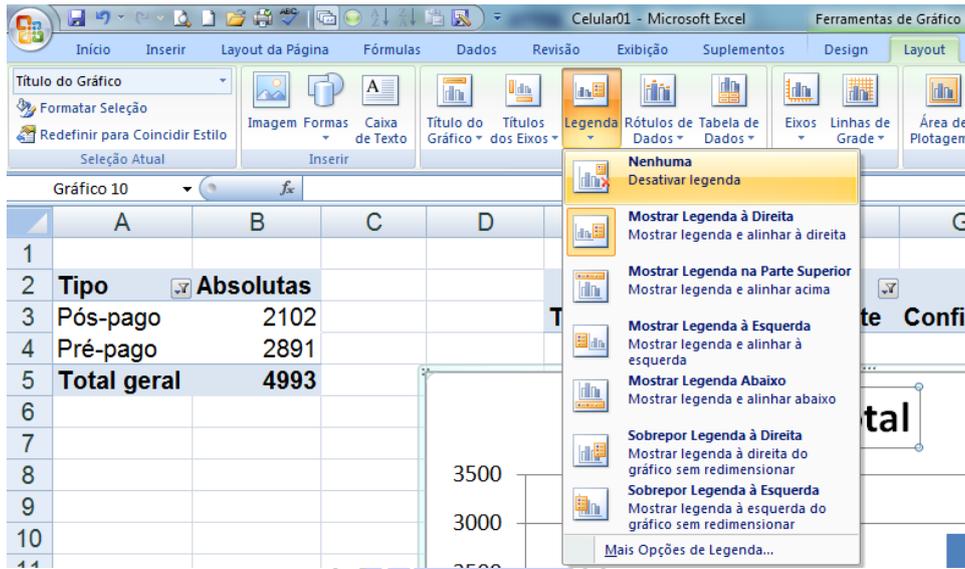


Figura 79 - Opções de legenda do gráfico dinâmico

Escolhendo “Nenhum” na Figura 79 a legenda desaparecerá. Mas, precisamos acrescentar um título ao eixo vertical, selecionando “Título dos Eixos” na Figura 78, e a opção “Título do Eixo Vertical Principal” chegamos à Figura 80.

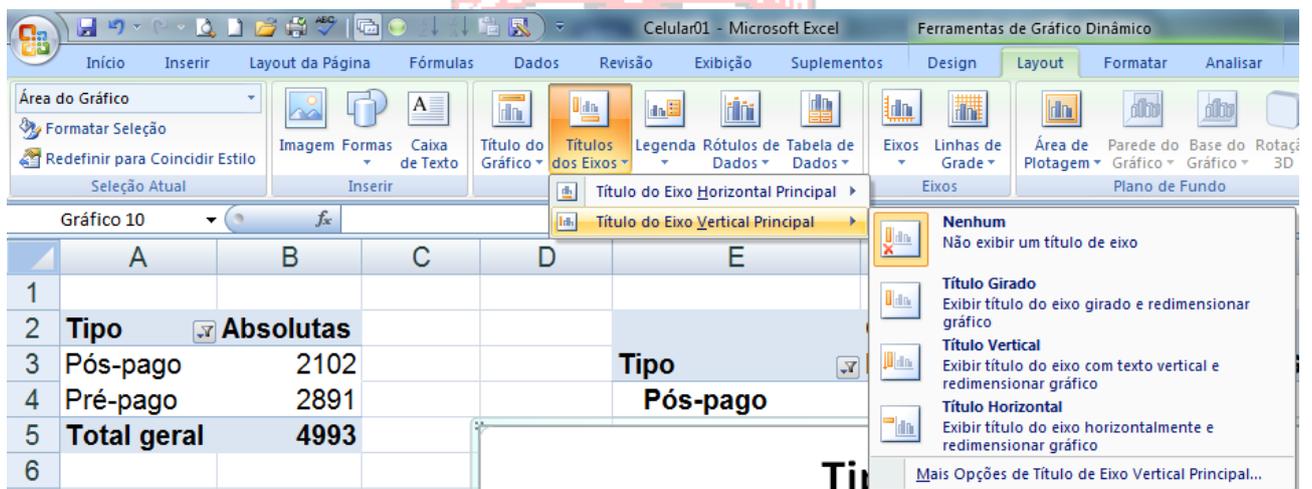


Figura 80 - Opções de Título do Eixo Vertical Principal em Gráfico Dinâmico

“Título Girado” permitirá que o nome acompanhe o eixo, e com esta seleção o Excel 2007 cria automaticamente uma caixa de texto chamada “Título do eixo”, que pode ser modificada exatamente como o título do gráfico. O resultado final, após a remoção da legenda e acréscimo dos títulos está na Figura 81

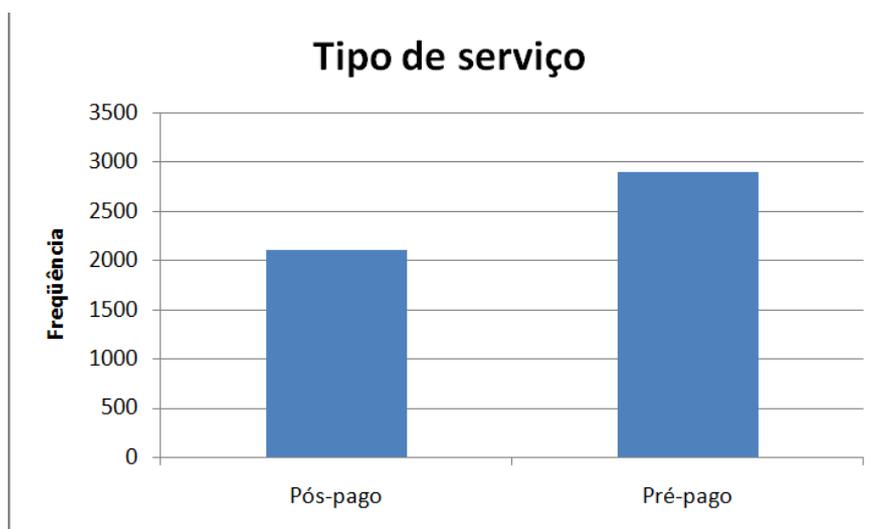


Figura 81 - Gráfico de colunas de Tipo de serviço - Final

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 82.

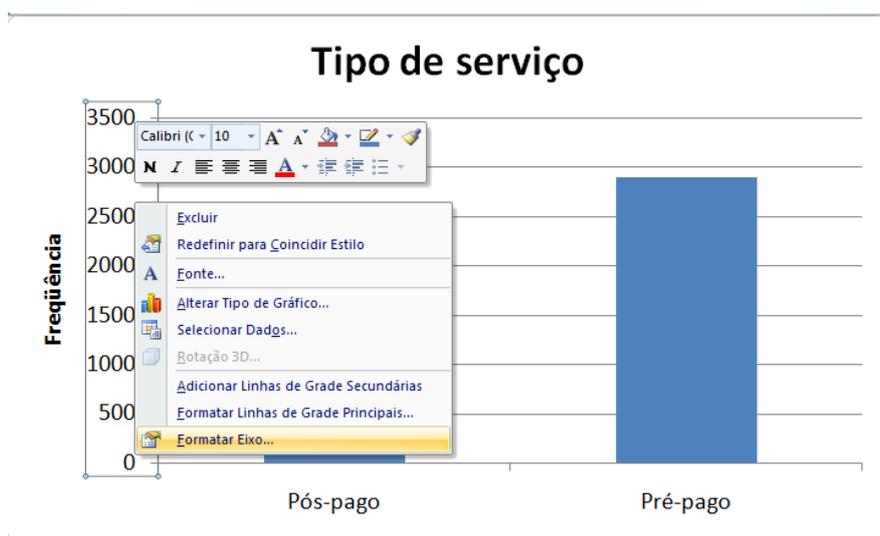


Figura 82 - Opções de formatação de gráfico e eixo

Pressionando “Formatar Eixo” nos leva à Figura 83. Automaticamente o Excel 2007 escolhe a escala mais apropriada de acordo com os dados de origem (Opções de Eixo: Automático). Mas, podemos trocar os limites e as unidades da escala. Na Figura 84 trocamos o mínimo para Fixo, indicando que o eixo vertical começará em 1000. O gráfico resultante é mostrado na Figura 85.

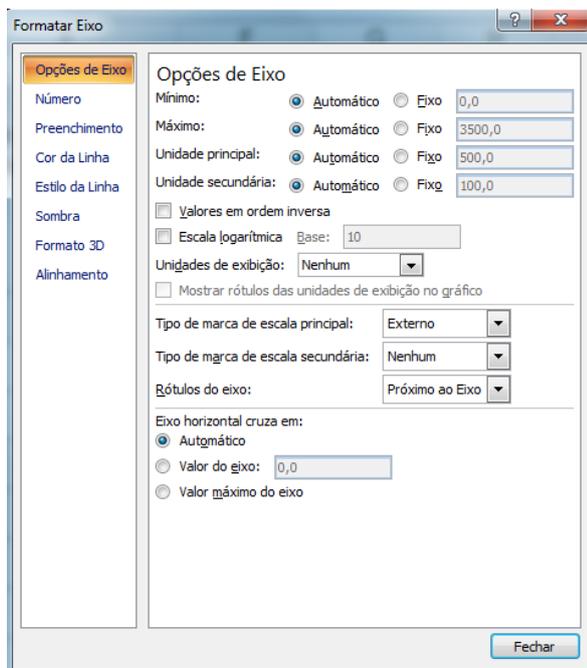


Figura 83 - Opções de Eixo

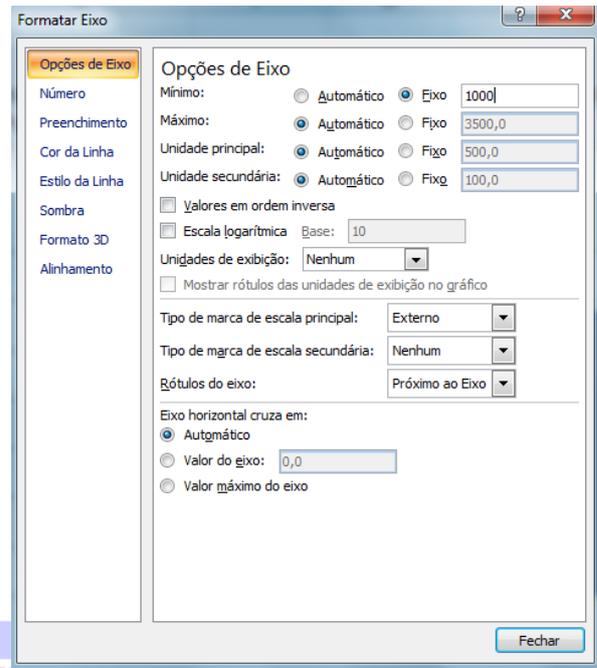


Figura 84 - Mudança de escala no eixo vertical

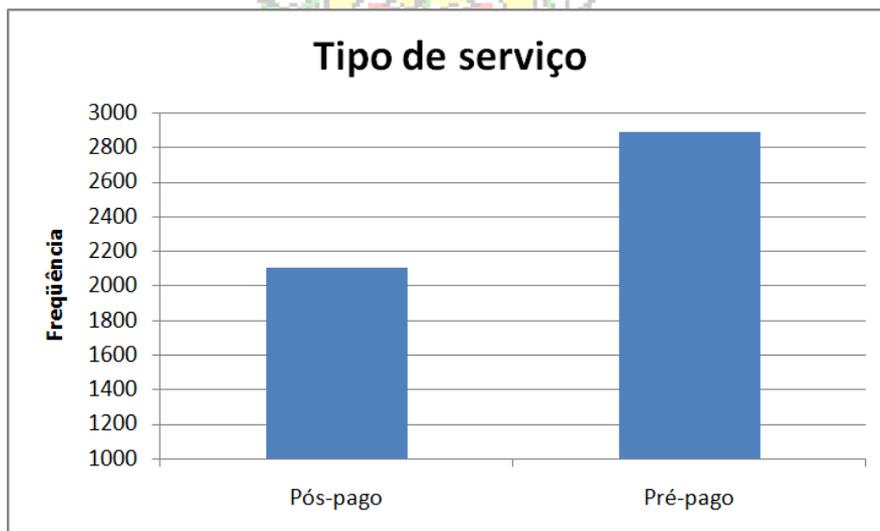


Figura 85 - Gráfico de colunas de Tipo de serviço – escala vertical modificada

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. A diferença é que ele apresentará automaticamente os percentuais de cada valor que ela pode assumir em relação ao total geral.

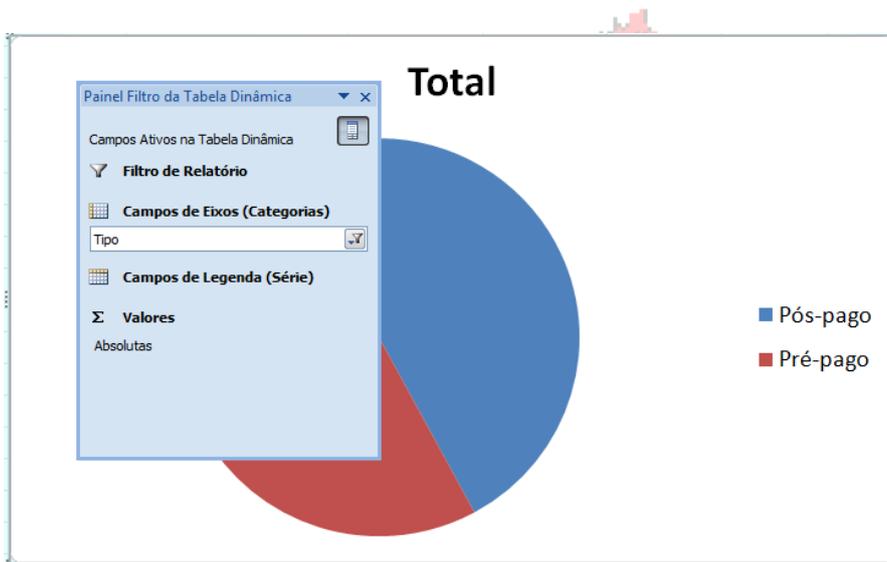
Ponha o cursor em algum ponto da tabela dinâmica de Tipo, e procure no menu Inserir por gráficos, como na Figura 71, mostrada anteriormente. Clique no tipo de gráfico e, em seguida, clique no subtipo de gráfico que deseja usar. Agora, queremos “Pizza”. Se pressionarmos a seta imediatamente abaixo de “Pizza” é possível ver todos os subtipos disponíveis (Figura 86).



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

Para a tabela dinâmica de Tipo vamos escolher Pizza 2D resultando na Figura 87.

Figura 86 - Subtipos de gráficos em Pizza



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

Figura 87 - Gráfico em pizza de Tipo de serviço - 1a versão

Para acrescentar os percentuais é preciso selecionar o gráfico e observar novamente as “Ferramentas do Gráfico Dinâmico”. As opções de Design na Figura 88 (veja que são adaptadas ao gráfico de Pizza), e as opções de Layout na Figura 89, incluindo “Rótulos de Dados”.



Figura 88 - Opções de design de gráfico dinâmico em Pizza



Figura 89 - Opções de layout de gráfico dinâmico em Pizza

Pressionando na seta ao lado de Rótulo de Dados chegamos à Figura 90.

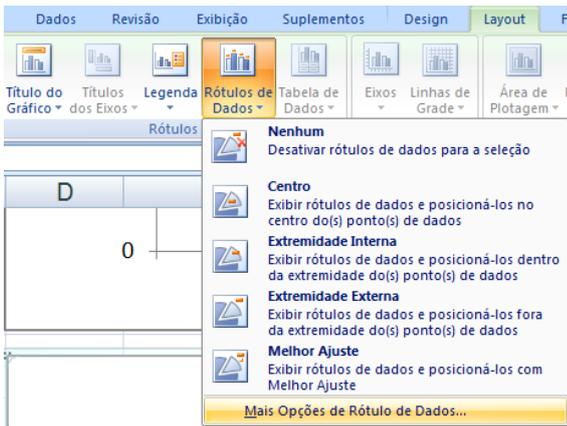


Figura 90 - Opções de Rótulo de Dados para Gráfico em Pizza

As opções se referem inicialmente à posição onde serão colocados os rótulos, destacando-se Centro e Melhor Ajuste.

Mas, precisamos de mais opções de Rótulos de Dados para acrescentar os percentuais de interesse. Pressionando “Mais Opções de Rótulos de Dados”, o resultado é a Figura 91. Em “Opções de Rótulo” é possível marcar Porcentagem e escolher sua posição como Melhor Ajuste, finalmente obtendo o gráfico da Figura 92.

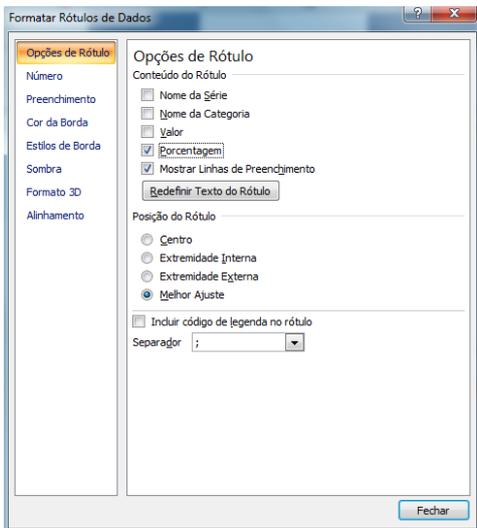


Figura 91 - Mais opções de Rótulo de Dados

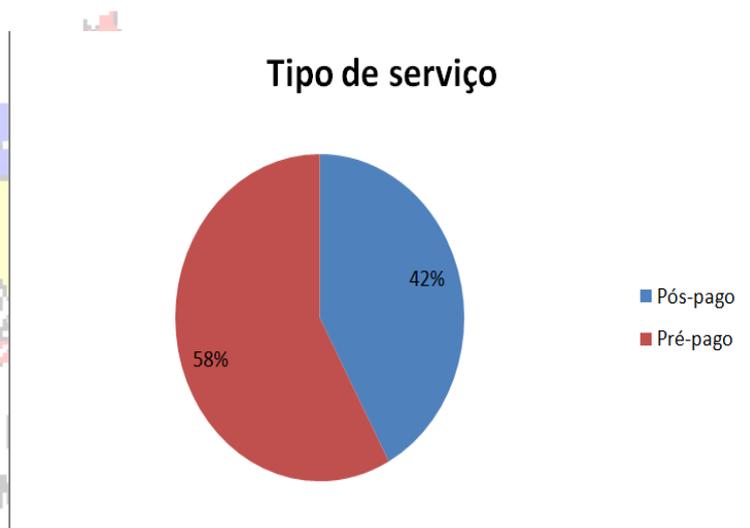


Figura 92 - Gráfico em pizza de Tipo de Serviço - Final

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 58. 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 93.

Frequência	Opinião					Total geral
Tipo	Excelente	Confiável	Mediana	Sofrível	Terrível	
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 93 - Tabela dinâmica de Tipo por Opinião - apenas frequências absolutas

Ponha o cursor em algum ponto da tabela dinâmica de Tipo por Opinião, e procure no menu “Inserir” por gráficos, como na Figura 71. Clique no tipo de gráfico e, em seguida, clique no subtipo de gráfico que deseja usar. No nosso caso, queremos “Colunas” (barras verticais). Se pressionarmos a seta imediatamente abaixo de “Colunas” é possível ver todos os subtipos disponíveis (Figura 72). Neste momento vamos escolher “Colunas 2D – agrupadas, o que resulta no gráfico da Figura 94.

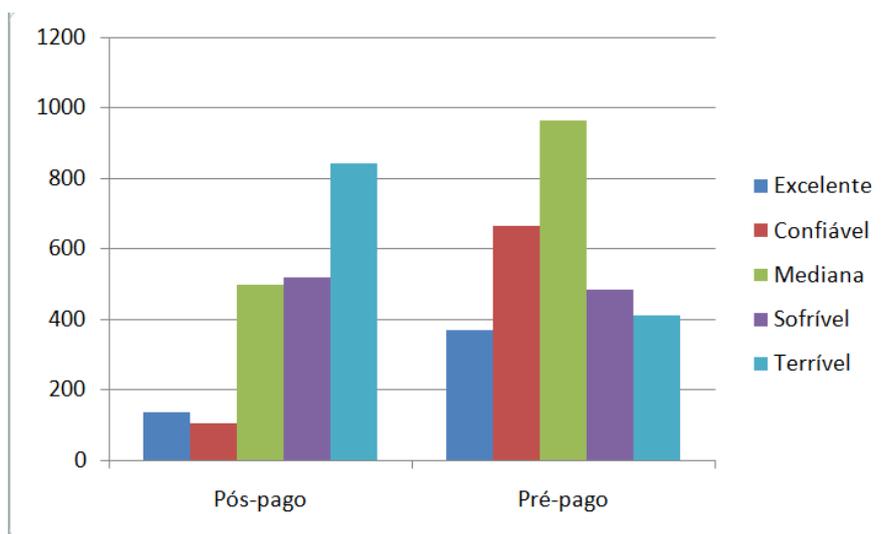


Figura 94 - Gráfico de colunas 2D agrupadas de Tipo por Opinião

Observe que não há problemas com os percentuais, nem com a escala que começa em zero. 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 95:

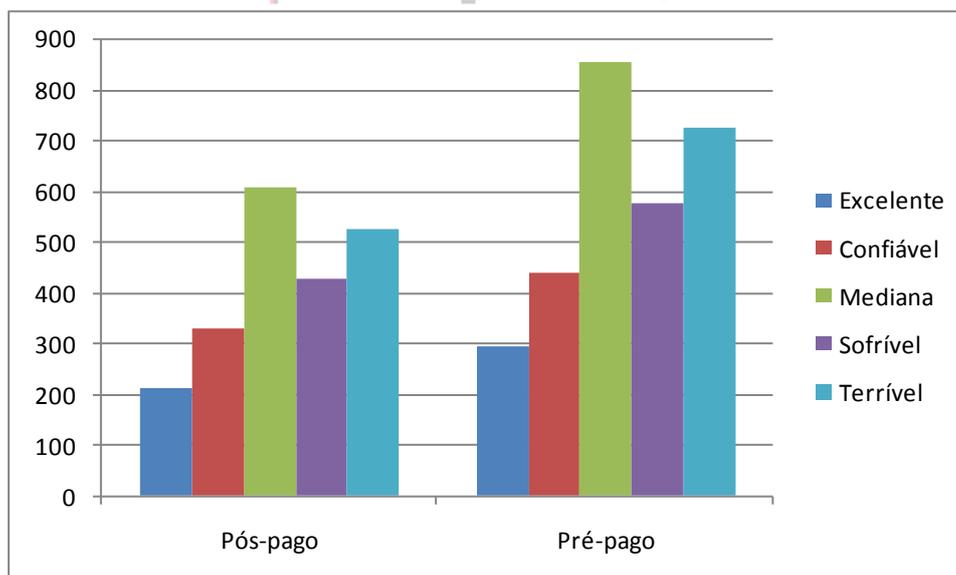


Figura 95 - Gráfico de colunas 2D agrupadas de Tipo por Opinião – Pouco relacionamento

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 94, basta selecionar o gráfico e pressionar o botão direito do mouse, o que faz surgir as opções da Figura 96.

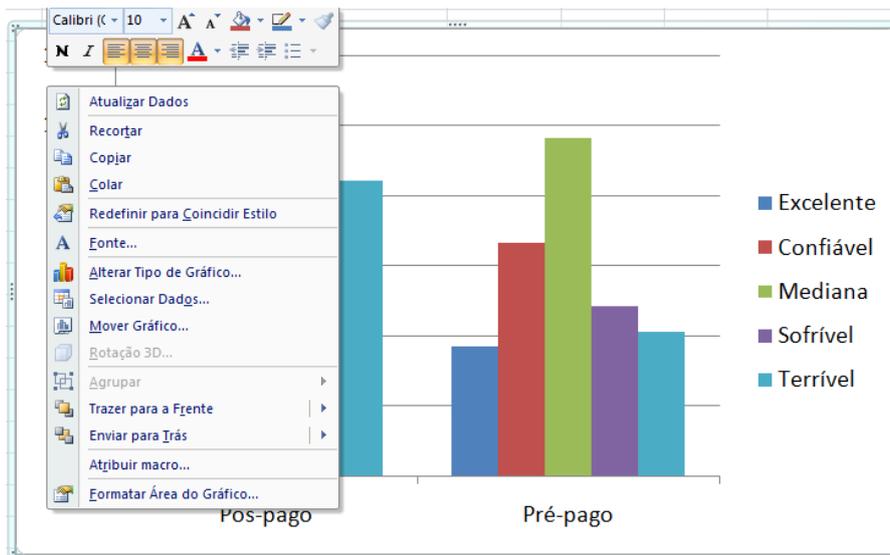


Figura 96 - Alteração do tipo de gráfico

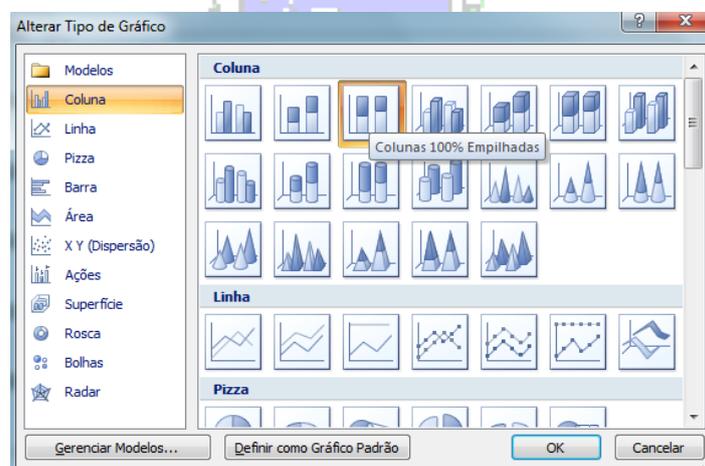


Figura 97 - Opções de alteração de tipo de gráfico

Como o gráfico de origem era de coluna seus subtipos aparecem em destaque, mas poderíamos escolher qualquer um. Mas, desejamos os percentuais, então escolhemos Colunas 2D 100% empilhadas, resultando na Figura 98.

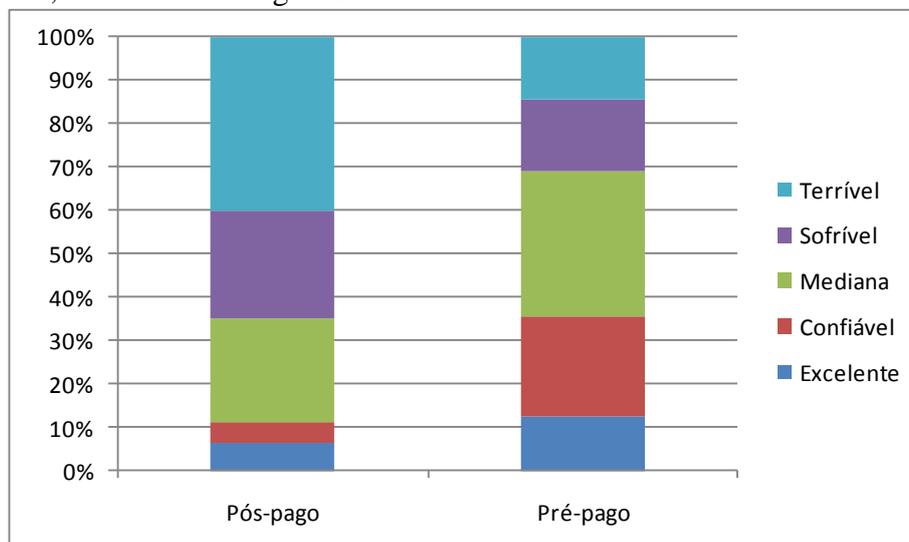


Figura 98 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião

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 58.

Ao modificar o gráfico da Figura 95 chegamos à Figura 99:

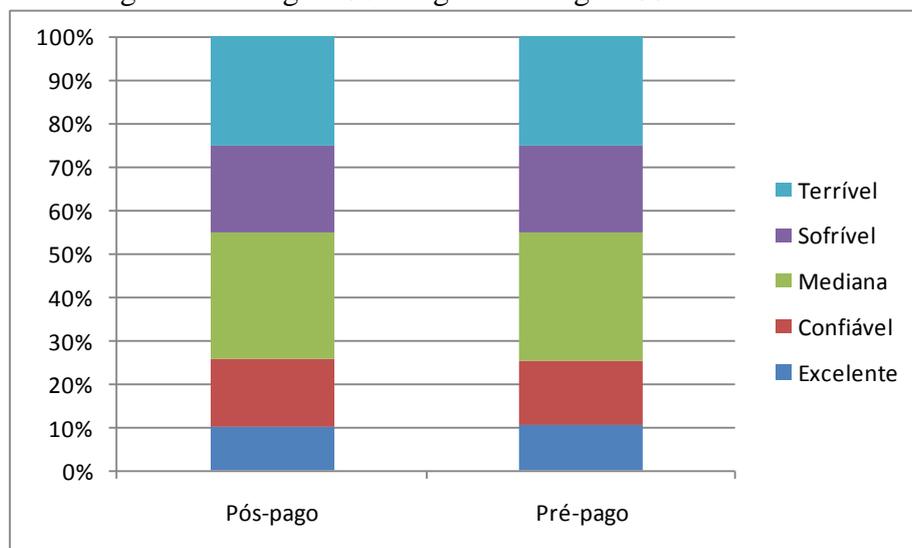


Figura 99 - Gráfico de colunas 2D 100% empilhadas de Tipo por Opinião – Pouco relacionamento

Ao contrário do caso da Figura 98 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 64) e outro para Masculino (Figura 65), e compará-los⁸.

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 37 a 40, pondo os resultados em uma planilha existente ou não. Chegamos então ao layout da tabela dinâmica de Anos, mostrado na Figura 100: a variável

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

Anos vai para Rótulos de Linha e Valores (não incluiremos nenhum percentual ainda). Após a inserção dos dados, e com pequenas modificações (incluindo a remoção das células vazias), surge a tabela da Figura 101.

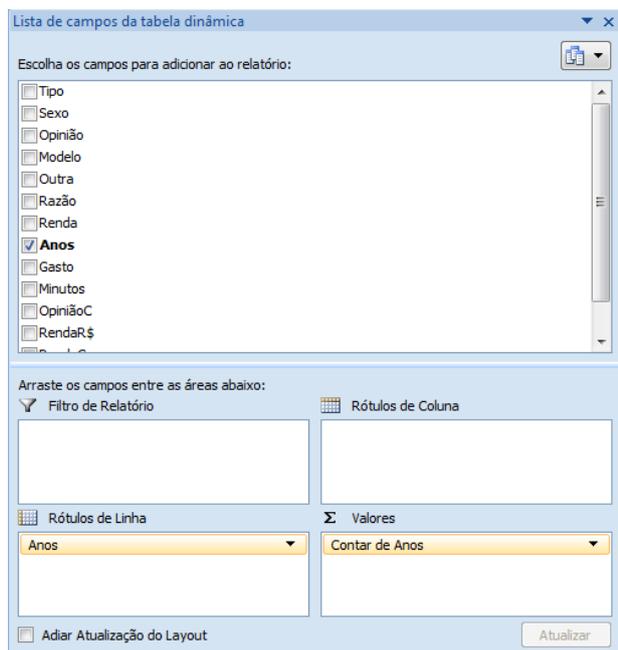


Figura 100 - Layout da tabela dinâmica de Anos

Anos	Freqüências
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 101 - Tabela dinâmica de Anos

A tabela da Figura 101 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 freqüência zero). Precisamos então fazer mais modificações para obter a tabela correta. Para tornar a tabela correta podemos copiá-la⁹ para células apropriadas em outra parte da planilha, incluindo as freqüências. Isso é mostrado na Figura 102.

Anos	Freqüências	Anos	Freqüências
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	Total geral	4996

Figura 102 - Cópia das células da Tabela dinâmica de Anos

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: selecionam-se as células onde queremos inserir as novas, no menu Início procuramos pela opção “Inserir”, e dentro desta escolhemos “Inserir células” como na Figura 103.

⁹ Ao fazer a colagem opte por “Colar valores”.

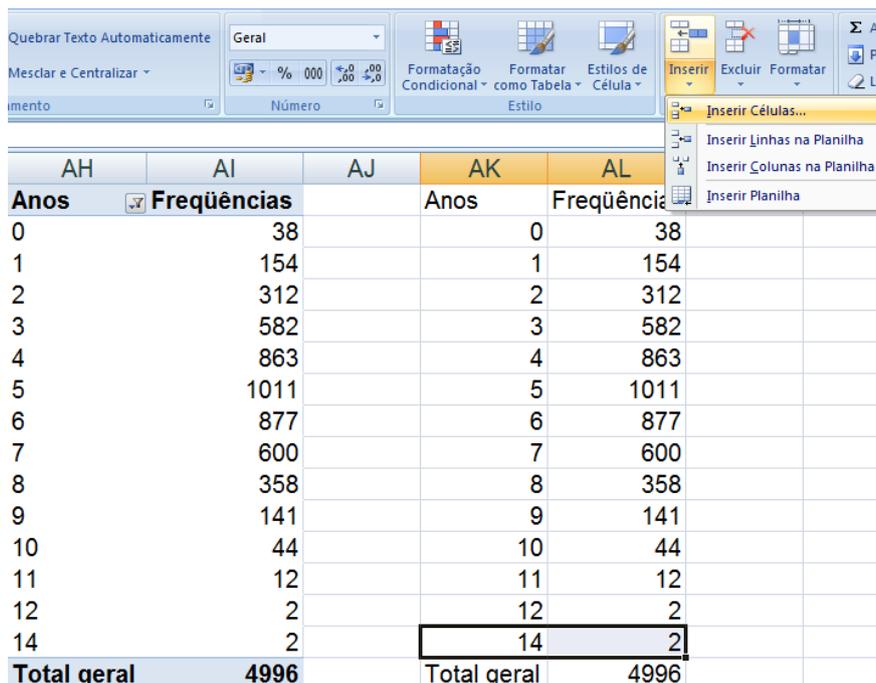


Figura 103 - Inserção de células no Excel 2007

Pressionando “Inserir células” o Excel apresentará a Figura 104. Basta escolher “Deslocar células para baixo”, resultando na Figura 105. 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 106).

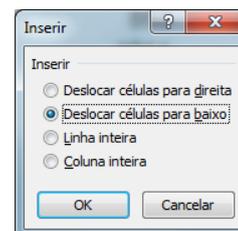


Figura 104 - Inserção de células: deslocamento para baixo

Anos	Frequências
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 105 - Células inseridas na Tabela de Anos

Anos	Frequências
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
13	0
14	2
Total geral	4996

Figura 106 - Tabela de frequências de Anos

Podemos acrescentar percentuais à tabela, através de regra de três simples, tendo 4996 como 100%, trocando o formato das células para %, o que resulta na Figura 107.

Anos	Freqüências	%
0	38	0,76%
1	154	3,08%
2	312	6,24%
3	582	11,65%
4	863	17,27%
5	1011	20,24%
6	877	17,55%
7	600	12,01%
8	358	7,17%
9	141	2,82%
10	44	0,88%
11	12	0,24%
12	2	0,04%
13	0	0,00%
14	2	0,04%
Total geral	4996	100,00%

Figura 107 - Tabela de Anos: freqüências e percentuais

5.1.1 – Distribuição de freqüência acumulada

Quando lidamos com variáveis quantitativas (discretas ou contínuas) torna-se interessante acrescentar à tabela as freqüências e percentuais acumulados, o que pode enriquecer a análise. Para tanto podemos acrescentar à tabela da Figura 107 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 dados¹⁰, este deverá ser o intervalo inserido na função CONT.SE. O critério deverá ser “igual ao conteúdo” da célula P2, que contém o primeiro valor da variável (Pós-pago). Vejam na Figura 108:

	AK	AL	AM	AN	AO	AP	AQ
1	Anos	Freqüências	%	Freq. Acumulada			
2	0	38	0,76%	=CONT.SE(Plan1!\$H\$2:\$H\$5001;"<="&Plan2!AK2)			
3	1	154	3,08%	CONT.SE(intervalo; critérios)			
4	2	312	6,24%				
5	3	582	11,65%				
6	4	863	17,27%				
7	5	1011	20,24%				
8	6	877	17,55%				
9	7	600	12,01%				
10	8	358	7,17%				
11	9	141	2,82%				
12	10	44	0,88%				
13	11	12	0,24%				
14	12	2	0,04%				
15	13	0	0,00%				
16	14	2	0,04%				
17	Total geral	4996	100,00%				

Figura 108 - CONT.SE: primeira freqüência acumulada da variável Ano

¹⁰ No arquivo Celular01.xlsx é a Plan1.

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 AK2 (de Plan2) 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 109.

	AK	AL	AM	AN
1	Anos	Freqüências	%	Freq. Acumulada
2	0	38	0,76%	38
3	1	154	3,08%	192
4	2	312	6,24%	504
5	3	582	11,65%	1086
6	4	863	17,27%	1949
7	5	1011	20,24%	2960
8	6	877	17,55%	3837
9	7	600	12,01%	4437
10	8	358	7,17%	4795
11	9	141	2,82%	4936
12	10	44	0,88%	4980
13	11	12	0,24%	4992
14	12	2	0,04%	4994
15	13	0	0,00%	4994
16	14	2	0,04%	4996
17	Total geral	4996	100,00%	

Pode-se perceber uma concentração dos clientes nos valores “centrais”, em torno de 5 anos. Podemos constatar:

- o valor 5 é o mais freqüente, 1011 pessoas são clientes há 5 anos completos;
- a esmagadora maioria das pessoas é cliente da operadora entre 3 a 7 anos (78,72%);
- clientes de “longo prazo” são relativamente poucos, pouco mais de 1% das pessoas está a 10 anos ou mais na operadora;
- a empresa parece não estar conseguindo aumentar sua base de clientes nos últimos tempos, apenas 10,90% das pessoas tornaram-se clientes nos últimos 3 anos.

Figura 109 - 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. Se pusermos o cursor em qualquer ponto da tabela da Figura 109 e selecionarmos Inserir, gráfico, colunas 2D, o resultado será a Figura 110.

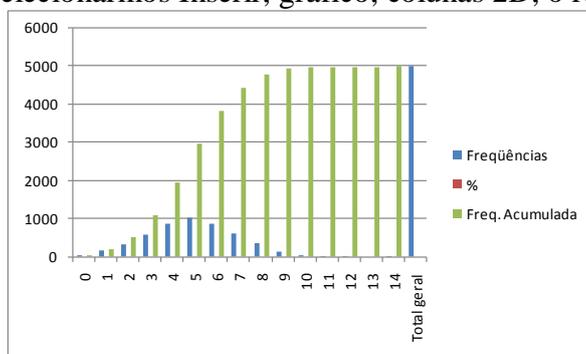


Figura 110 - Histograma de Anos - 1a versão

Observe que o gráfico tem várias séries que não nos interessam (% , freqüência acumulada, além do total geral). Todas estas precisam ser removidas: seleciona-se o gráfico e pressiona-se o botão direito do mouse, resultando na Figura 111.

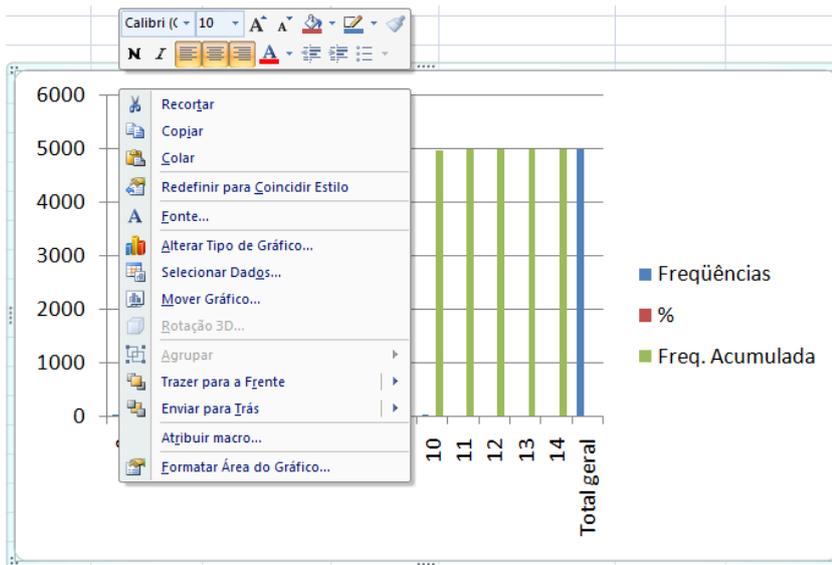


Figura 111 - Opções de gráfico - histograma de Anos

Para remover os dados desnecessários pressiona-se “Selecionar Dados”, o que nos leva à Figura 112. Marca-se % e pressiona-se “Remover”, também se pressiona “Remover”. Além disso, os dados incluem até a linha 17, o total geral. Portanto, em “Intervalo de dados do gráfico” devemos trocar o 17 por 16:

=Plan2!\$AK\$1:\$AN\$17 por
 =Plan2!\$AK\$1:\$AN\$16
 Resultado na Figura 113.

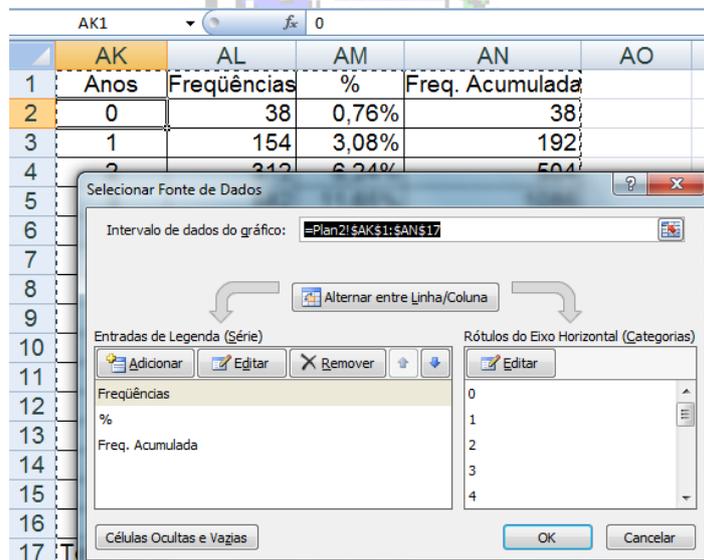


Figura 112 - Seleção de fontes de dados - histograma de Anos

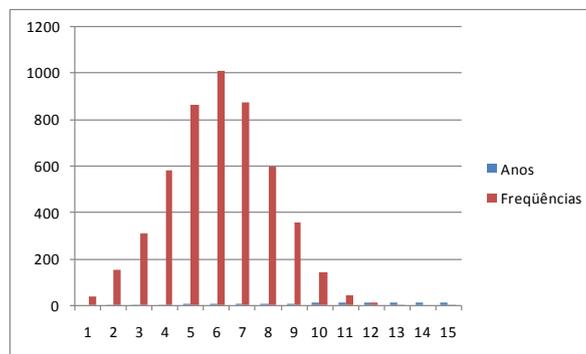
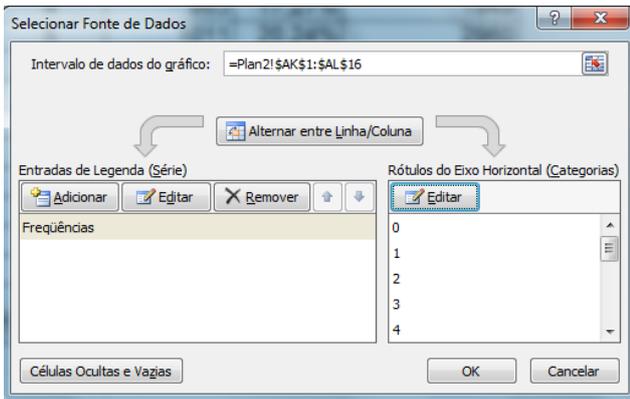


Figura 113 - Histograma de Anos - 2a versão

Há um problema com o histograma da Figura 113: os dados dos Anos, que eram rótulos do eixo X “moveram-se” para o eixo Y, o que não está correto. Novamente, precisamos selecionar os dados, como na Figura 111, levando à Figura 114.



A série “Anos” foi removida das “Entradas de Legenda” e no campo “Rótulos do Eixo Horizontal”, pressionamos “Editar” e selecionamos o intervalo Plan2!AK2:AK16 que contém os valores dos Anos descritos na tabela da Figura 109.

Pressionando OK na Figura 114 chegamos ao histograma da Figura 115.

Figura 114 - Seleção de fontes de dados - histograma de Anos - modificações

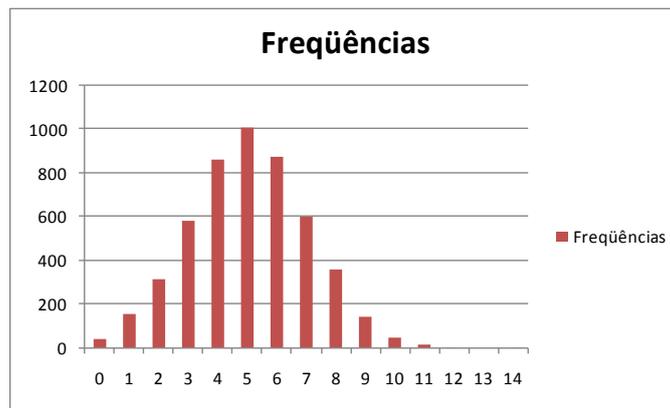


Figura 115 - Histograma de Anos - 3a versão

Podemos mudar o título, remover a legenda (já que há apenas uma série de dados), entre outras alterações cosméticas. Mas, foi mencionado que as colunas do histograma precisam ser justapostas. Portanto, precisamos reduzir o espaçamento entre elas a zero. Basta selecionar as colunas e pressionar o botão direito do mouse (Figura 116).

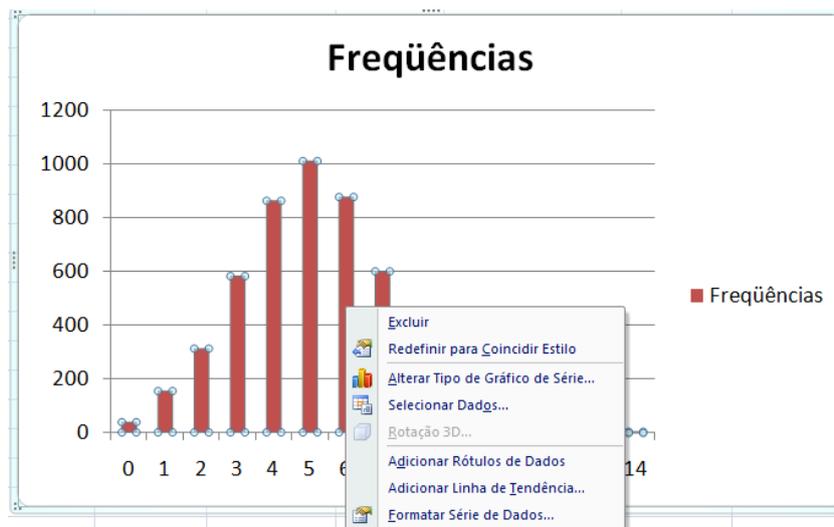


Figura 116 - Opções de formatação de séries de dados - histograma Anos

Escolhendo “Formatar Série de Dados” surge a Figura 117.

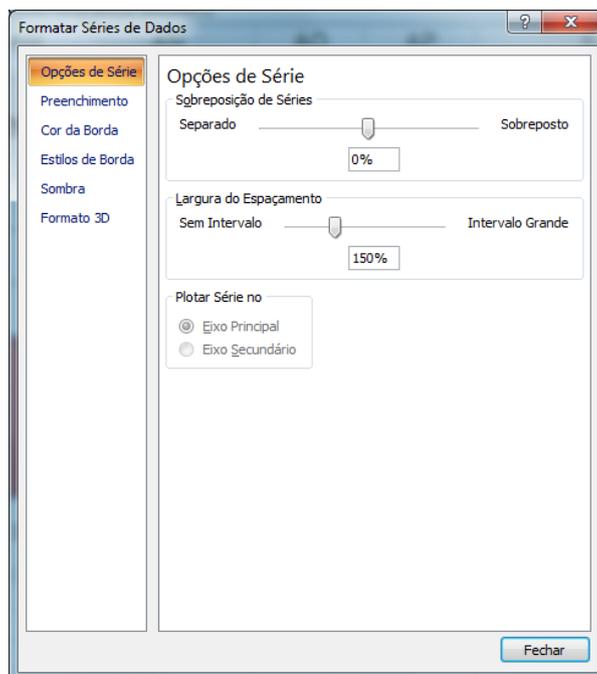


Figura 117 - Opções de formatar séries de dados

Reduzindo o campo “Largura do espaçamento” a Sem intervalo, além das mudanças nos títulos e legenda, o histograma correto é mostrado na Figura 118.

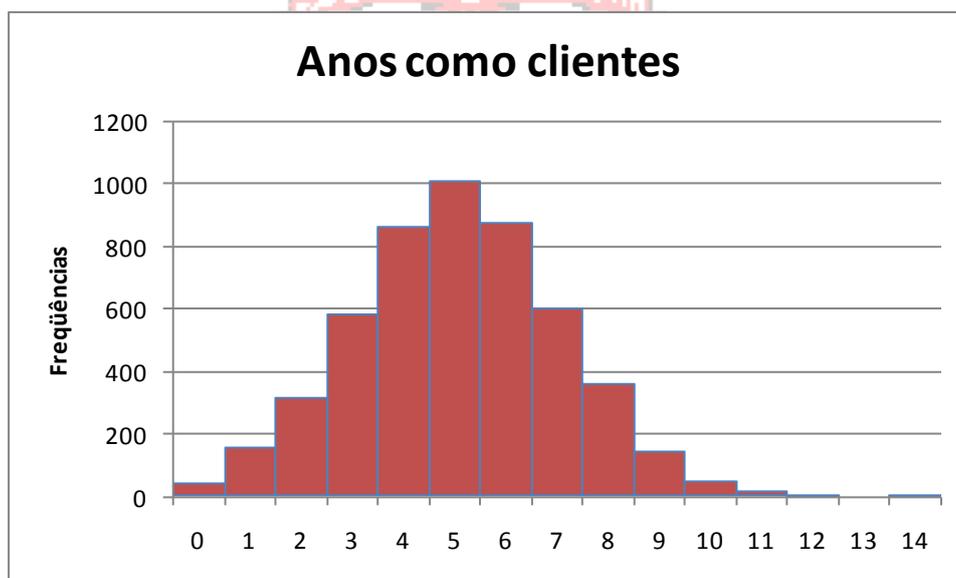


Figura 118 - Histograma de Anos completos como cliente

Percebe-se claramente que os anos estão concentrados em torno de 5 anos, 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 AQ2 e o máximo na célula AQ3, todas na Plan2. Selecione esta célula com o cursor. Podemos inserir as funções ou escrevê-las diretamente nas células:

= MÍNIMO(Plan1!G2:G5001) = MÁXIMO(Plan1!G2:G5001) Veja a Figura 119.

	AQ	AR	AS	AT		AQ	AR	AS	AT
1	Renda				1	Renda			
2	Mínimo	=MÍNIMO(Plan1!G2:G5001)			2	Mínimo	1,003		
3	Máximo				3	Máximo	=MÁXIMO(Plan1!G2:G5001)		

Figura 119 - Mínimo e máximo para a variável Renda

Para calcular o intervalo basta colocar a fórmula respectiva em uma célula, AR4 por exemplo. Veja a Figura 120.

	AQ	AR
1	Renda	
2	Mínimo	1,003
3	Máximo	43,878
4	Intervalo	=AR3-AR2

Observa-se, conforme anteriormente 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.

Figura 120 - Intervalo para Renda

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, extraíndo a raiz quadrada (a função RAIZ, na categoria "Matemática e trigonométrica" 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 121.

	AQ	AR
1	Renda	
2	Mínimo	1,003
3	Máximo	43,878
4	Intervalo	42,875
5	Número de classes	10
6	Amplitude das classes	=AR4/AR5

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, com 10 classes, o resultado poderá ser calculado em células apropriadas, como mostrado a partir da Figura 122.

Figura 121 - Amplitude de classes de Renda

	AQ	AR		AQ	AR
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 de classes	10	5	Número de classes	10
6	Amplitude das classes	4,2875	6	Amplitude das classes	4,2875
7			7		
8	Classes		8	Classes	
9	Limite inferior	Limite superior	9	Limite inferior	Limite superior
10	=AR2		10	1,003	=AQ10+\$AR\$6

Figura 122 – Cálculo dos limites inferior e do limite superior da primeira classe de Renda

O limite inferior da primeira classe será o conteúdo da célula AR2 (mínimo de Renda), e o limite superior será o limite inferior (célula AR10) mais o conteúdo da célula AR6 (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 123.

	AQ	AR		AQ	AR
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 de classes	10	5	Número de classes	10
6	Amplitude das classes	4,2875	6	Amplitude das 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	=AR10		11	5,2905	=AQ11+\$AR\$6

Figura 123 – 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 AQ11 e AR11 até o fim da tabela e completar as 10 classes. Veja a Figura 124.

	AQ	AR
1	Renda	
2	Mínimo	1,003
3	Máximo	43,878
4	Intervalo	42,875
5	Número de classes	10
6	Amplitude das 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 124 - Classes de Renda

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

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

Precisamos também definir os pontos médios das classes, que passarão a representá-las, como se as frequências obtidas fossem deles. Esses procedimentos são mostrados nas Figuras 125 a 127.

	AQ	AR	AS	AT	AU
1	Renda				
2	Mínimo	1,003			
3	Máximo	43,878			
4	Intervalo	42,875			
5	Número de classes	10			
6	Amplitude das classes	4,2875			
7	Classes				
9	Limite inferior	Limite superior	Ponto Médio	Freqüências	%
10	1,003	5,2905	$=(AQ10+AR10)/2$		
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 125 - Ponto médio da primeira classe de Renda

Basta arrastar a fórmula da célula AS10 até a AS19 para obter os pontos médios das classes.

	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB
1	Renda											
2	Mínimo	1,003										
3	Máximo	43,878										
4	Intervalo	42,875										
5	Número de classes	10										
6	Amplitude das classes	4,2875										
7	Classes											
9	Limite inferior	Limite superior	Ponto Médio	Freqüências	%							
10	1,003	5,2905	3,14675	$=CONT.SE(Plan1!$G$2:$G$5001;"<"&Plan2!AR10)-CONT.SE(Plan1!G2:G5001;"<"&Plan2!AQ10)$								
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 126 - Obtenção da freqüência da primeira classe de Renda

Observe que se contam as ocorrências menores do que o limite superior da classe, subtraídas da contagem das ocorrências menores do que o limite inferior, mantendo o intervalo G2 a G5001 (da Plan1- valores de Renda) como referência absoluta, para permitir arrastar a fórmula até a célula AT19.

	AQ	AR	AS	AT	AU	AV	AW	AX	AY	AZ	BA	BB	BC
1	Renda												
2	Mínimo	1,003											
3	Máximo	43,878											
4	Intervalo	42,875											
5	Número de classes	10											
6	Amplitude das classes	4,2875											
7													
8	Classes												
9	Limite inferior	Limite superior	Ponto Médio	Freqüências	%								
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										

Figura 127 - Obtenção da freqüência da última classe de Renda

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

Podemos também obter freqüências e percentuais acumulados para a tabela de Renda. Basta remover o segundo membro das fórmulas mostradas nas Figuras 126 e 127, tomando os mesmos cuidados com o limite superior da última classe, como mostrado nas Figuras 128 e 129.

	AQ	AR	AS	AT	AU	AV	AW	AX	AY
1	Renda								
2	Mínimo	1,003							
3	Máximo	43,878							
4	Intervalo	42,875							
5	Número de classes	10							
6	Amplitude das classes	4,2875							
7									
8	Classes								
9	Limite inferior	Limite superior	Ponto Médio	Freqüências	%	Freq.acumulada	% acumulado		
10	1,003	5,2905	3,14675	2959	59,23%	=CONT.SE(Plan1!\$G\$2:\$G\$5001;"<"&Plan2!AR10)			
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%				

Figura 128 - Obtenção da freqüência acumulada da primeira classe de Renda

	AQ	AR	AS	AT	AU	AV	AW	AX	AY
1	Renda								
2	Mínimo	1,003							
3	Máximo	43,878							
4	Intervalo	42,875							
5	Número de classes	10							
6	Amplitude das classes	4,2875							
7									
8	Classes								
9	Limite inferior	Limite superior	Ponto Médio	Freqüências	%	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%	4996			

Figura 129 - Obtenção da frequência 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 130.

	AQ	AR	AS	AT	AU	AV	AW
1	Renda						
2	Mínimo	1,003					
3	Máximo	43,878					
4	Intervalo	42,875					
5	Número de classes	10					
6	Amplitude das classes	4,2875					
7							
8	Classes						
9	Limite inferior	Limite superior	Ponto Médio	Freqüências	%	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 130 - Renda agrupada em classes: frequências absolutas, percentuais, frequências acumuladas e percentuais acumulados

A maioria absoluta (na realidade, mais, pois são 59,23%) dos clientes tem 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

De forma análoga ao caso das variáveis quantitativas discretas podemos construir um histograma para uma variável quantitativa contínua, desde que ela já esteja resumida em uma tabela agrupada em classes. Agora iremos usar um procedimento diferente de inserção de gráfico, para evitar a necessidade de tantas correções como no caso do histograma de Anos. Devemos pôr o cursor em uma célula *vazia* da planilha e no menu “Inserir” escolher Gráfico, Colunas 2D, colunas agrupadas. Surgirá um gráfico totalmente em branco. Selecionando este gráfico e pressionando o botão direito do mouse vamos obter as opções de gráfico vistas na Figura 111. Escolhendo “Selecionar dados” poderemos selecionar para “Entradas de legenda” as frequências da tabela mostrada na Figura 130 (células AT10 a AT19 da Plan2), e para “Rótulos do eixo horizontal” as células que contém os limites inferiores e superiores das classes, como mostrado na Figura 131.

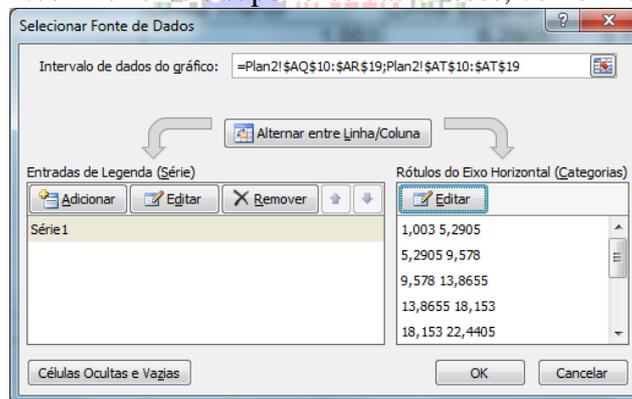


Figura 131 - Seleção de dados para Histograma de Renda

Pressionando OK o histograma resultante pode ser visto na Figura 132. Novamente, é preciso acrescentar título geral, títulos nos eixos, e reduzir o espaçamento entre as barras. Após todas essas operações o histograma correto está na Figura 132.



Figura 132 – 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 130, 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 119). Ou, podemos usar os menus do Excel 2007: basta selecionar “Fórmulas” e surge a Figura 133.

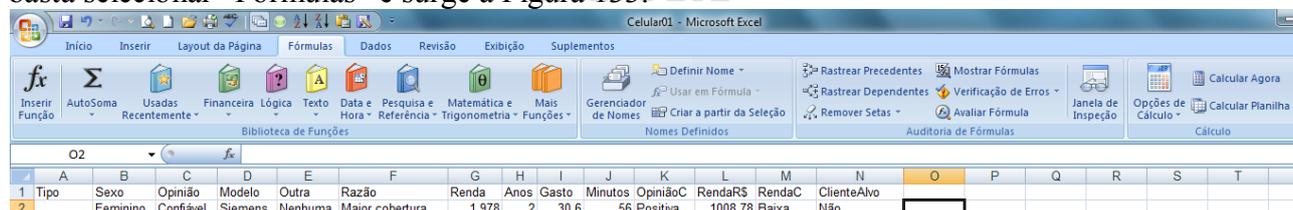


Figura 133 - Fórmulas no Excel 2007

Na extrema esquerda da Figura 133 está a opção “Inserir Função”¹¹. Observe também as opções: “Usadas Recentemente”, “Financeira”, etc. E, ao final, “Mais Funções”. Se pressionarmos esta última chegamos à Figura 134.

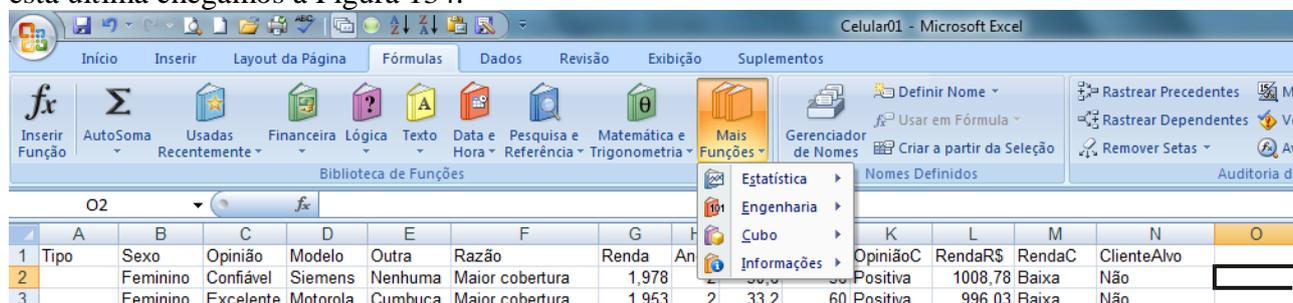
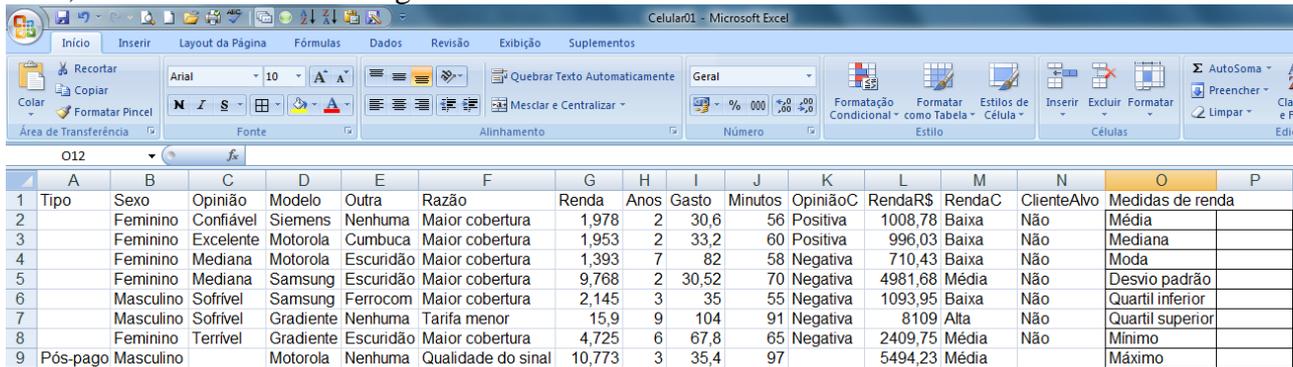


Figura 134 - Fórmulas no Excel 2007 – Mais funções

Dentre as opções vemos “Estatística”, que inclui as medidas de síntese que procuramos. Podemos inserir as medidas usando Inserir Função (Figura 133) ou pelo caminho da Figura 134. Mas, antes de tudo, devemos definir as células da Plan1 que armazenarão os resultados. Por conveniência os

¹¹ Exatamente como nas versões anteriores do Excel.

resultados das medidas serão postos nas células P2 a P8, com os nomes das medidas nas células O2 a O8, tal como mostrado na Figura 135.

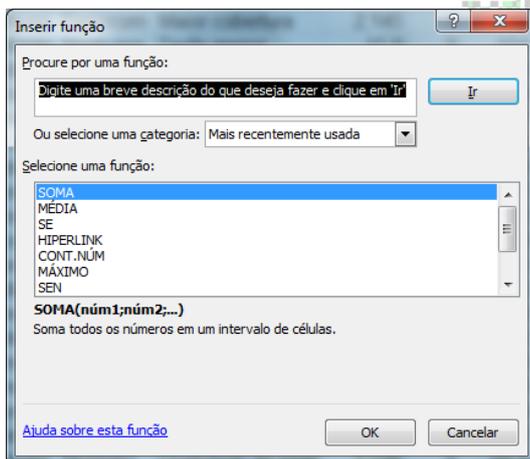


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Tipo	Sexo	Opinião	Modelo	Outra	Razão	Renda	Anos	Gasto	Minutos	OpiniãoC	RendaR\$	RendaC	ClienteAvo	Medidas de renda	
2		Feminino	Confiável	Siemens	Nenhuma	Maior cobertura	1,978	2	30,6	56	Positiva	1008,78	Baixa	Não	Média	
3		Feminino	Excelente	Motorola	Cumbuca	Maior cobertura	1,953	2	33,2	60	Positiva	996,03	Baixa	Não	Mediana	
4		Feminino	Mediana	Motorola	Escuridão	Maior cobertura	1,393	7	82	58	Negativa	710,43	Baixa	Não	Moda	
5		Feminino	Mediana	Samsung	Escuridão	Maior cobertura	9,768	2	30,52	70	Negativa	4981,68	Média	Não	Desvio padrão	
6		Masculino	Sofrível	Samsung	Ferrocom	Maior cobertura	2,145	3	35	55	Negativa	1093,95	Baixa	Não	Quartil inferior	
7		Masculino	Sofrível	Gradiente	Nenhuma	Tarifa menor	15,9	9	104	91	Negativa	8109	Alta	Não	Quartil superior	
8		Feminino	Terrível	Gradiente	Escuridão	Maior cobertura	4,725	6	67,8	65	Negativa	2409,75	Média	Não	Mínimo	
9	Pós-pago	Masculino		Motorola	Nenhuma	Qualidade do sinal	10,773	3	35,4	97		5494,23	Média		Máximo	

Figura 135 - Medidas de síntese de Renda - 1a parte

Medidas de síntese pelo botão Inserir Função

Vamos começar com a média. Posicione o cursor na célula P2. Ao pressionar o botão “Inserir Função” surge a tela da Figura 136.



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 137), entre elas a Estatística. Ao escolhermos esta todas as funções existentes desta categoria serão apresentadas, bastando percorrer a barra de rolagem para encontrar as que procuramos, como, por exemplo, a média.

Figura 136 - Funções disponíveis no Excel

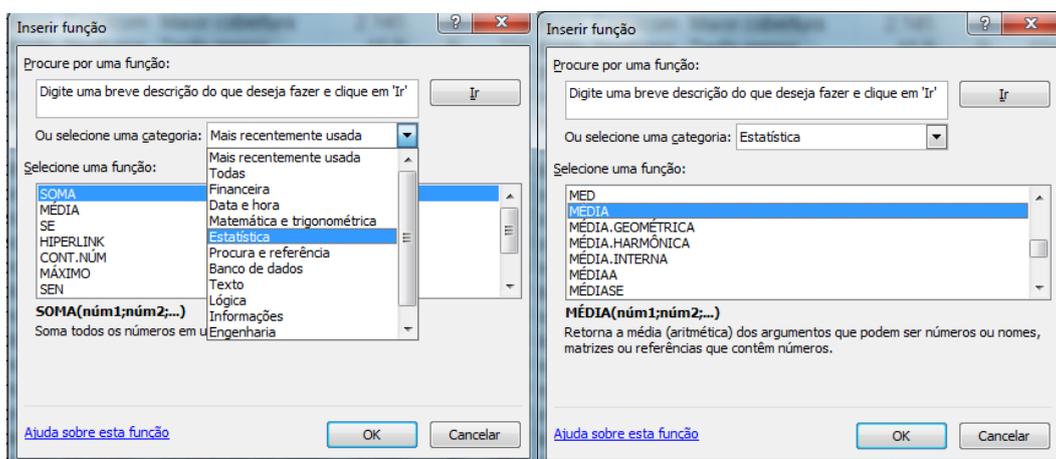


Figura 137 - Funções estatísticas no Excel - 1a abordagem

Para cada função o Excel apresenta a sua sintaxe (os argumentos necessários) e uma explicação sobre o que ela faz. Se pressionarmos OK passamos à tela da Figura 138.

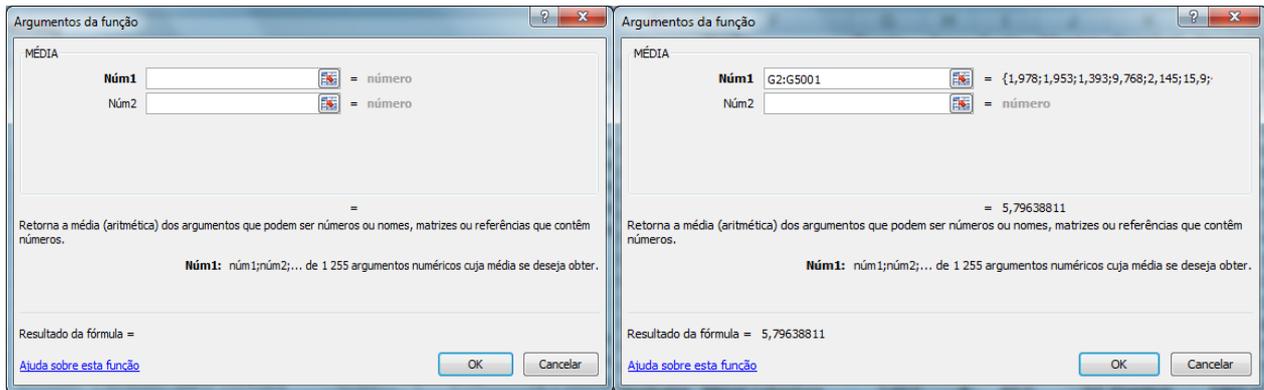


Figura 138 - Argumentos da função Média (para Renda) - 1a versão

Os dados de Renda estão nas células G2 a G5001¹². 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, 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, vamos completar os cálculos, os procedimentos estão nas Figuras 139 a 142.

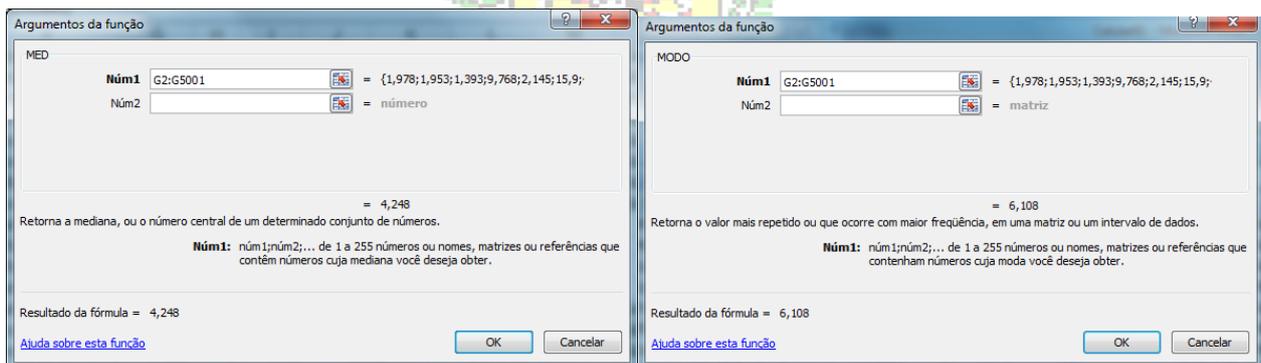


Figura 139 - Argumentos das funções Mediana e Moda (Renda) - 1a versão

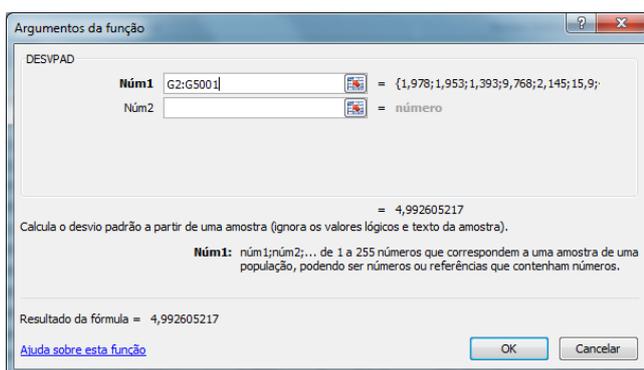


Figura 140 - Argumentos da função desvio padrão (Renda) - 1a versão

É importante ressaltar que o arquivo Celular01.xlsx refere-se a uma AMOSTRA, então devemos calcular o desvio padrão amostral com $n-1$ no denominador da expressão. No Excel isso é feito através da função `DESPVAD()`. Se houver interesse no desvio padrão populacional, com n no denominador, deve-se usar a função `DESPVADP`.

¹² A célula G1 é excluída por conter o nome da variável.

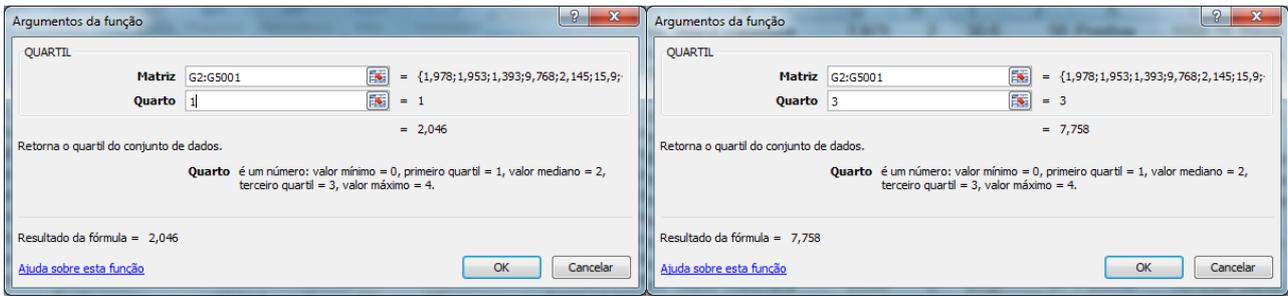


Figura 141 - Argumentos das funções Quartil Inferior e Superior (Renda) - 1a versão

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

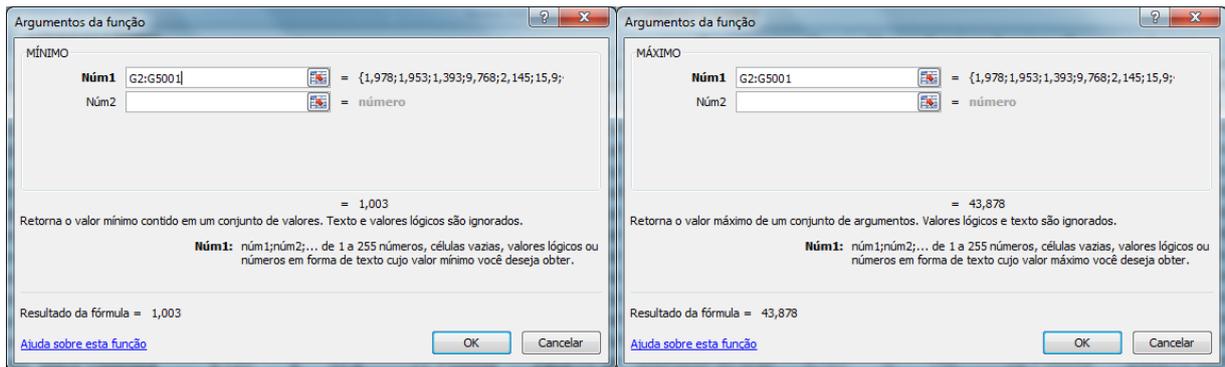


Figura 142 - Argumentos das funções Mínimo e Máximo (Renda) - 1a versão

E os resultados completos podem ser vistos na Figura 143.

	O	P
1	Medidas de renda	
2	Média	5,79639
3	Mediana	4,248
4	Moda	6,108
5	Desvio padrão	4,99261
6	Quartil inferior	2,046
7	Quartil superior	7,758
8	Mínimo	1,003
9	Máximo	43,878

Figura 143 - 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 143: coeficiente de variação percentual, diferenças entre os quartis e a mediana, e os valores que definem os valores discrepantes. Nas Figuras 144 a 146 apresentamos as operações:

	O	P	Q	R
1	Medidas de renda			
2	Média	5,79639	CV%	$= (P5/P2) * 100$
3	Mediana	4,248		
4	Moda	6,108		
5	Desvio padrão	4,99261		
6	Quartil inferior	2,046		
7	Quartil superior	7,758		
8	Mínimo	1,003		
9	Máximo	43,878		

Figura 144 - Cálculo do CV%

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.

	O	P	Q	R
1	Medidas de renda			
2	Média	5,79639	CV%	86,133
3	Mediana	4,248	Qs-Md	=P7-P3
4	Moda	6,108		
5	Desvio padrão	4,99261		
6	Quartil inferior	2,046		
7	Quartil superior	7,758		
8	Mínimo	1,003		
9	Máximo	43,878		

Figura 145 - 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).

	O	P	Q	R	S
1	Medidas de renda				
2	Média	5,79639	CV%	86,133	
3	Mediana	4,248	Qs-Md	3,51	
4	Moda	6,108	Md-Qi	2,202	
5	Desvio padrão	4,99261	Qi-1,5*(Qs-Qi)	=P6-1,5*(P7-P6)	
6	Quartil inferior	2,046			
7	Quartil superior	7,758			
8	Mínimo	1,003			
9	Máximo	43,878			

Figura 146 - 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 147.

	O	P	Q	R
1	Medidas de renda			
2	Média	5,79639	CV%	86,133
3	Mediana	4,248	Qs-Md	3,51
4	Moda	6,108	Md-Qi	2,202
5	Desvio padrão	4,99261	Qi-1,5*(Qs-Qi)	-6,522
6	Quartil inferior	2,046	Qs-1,5*(Qs-Qi)	16,326
7	Quartil superior	7,758		
8	Mínimo	1,003		
9	Máximo	43,878		

Figura 147 - Medidas de síntese para Renda (completas)

E a descrição da variável Renda será mostrada a seguir.

Tendência Central:

- o valor típico de Renda oscila entre 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 Brasil¹³.
- a moda vale 6,108 salários mínimos, mas para variáveis quantitativas contínuas esta medida não tem muito significado, pois os valores se repetem pouco¹⁴.

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;

¹³ É FUNDAMENTAL ter algum padrão para comparação, o que foi possível neste caso.

¹⁴ Usando a função CONT.SE descobre-se que a moda ocorreu apenas 31 vezes em um total de 5000 observações.

- 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) – valores de 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, o que é impossível;
- há discrepantes superiores, pois há valores acima¹⁵ 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 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 19 a 23). 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 138 a 142 trocando o intervalo de células lá utilizado pelos mencionados acima, resultando na Figura 148:

¹⁵ Usando a função CONT.SE verifica-se que há 224 clientes com renda acima deste valor, sendo portanto valores discrepantes.

	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 148 - 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).

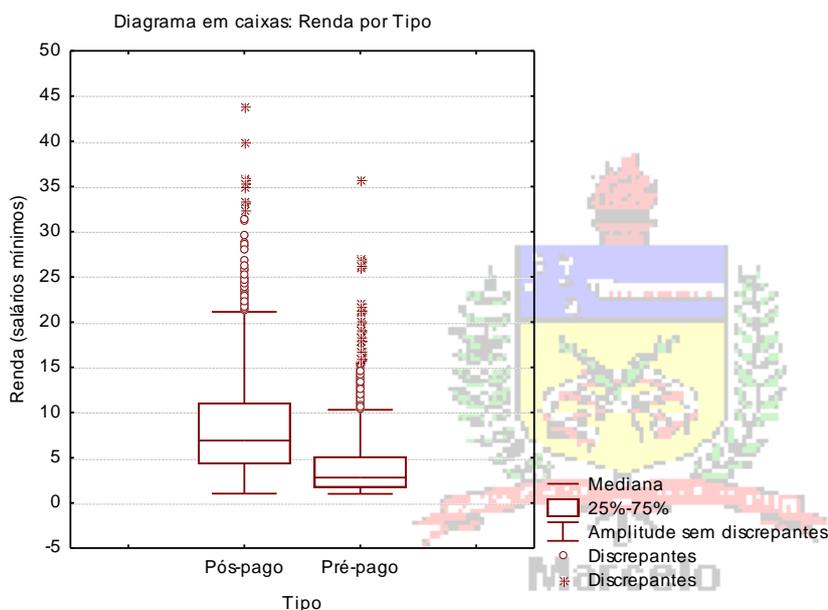
- média e mediana também são diferentes nos dois Tipos, constatando-se também pelas diferenças entre quartis e mediana;

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-pago¹⁶.

As diferenças entre as Rendas dos clientes de Pós-pago e Pré-pago podem ser detectadas em um diagrama em caixas múltiplo (Figura 149):



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 ($Q_3 - Q_1$) maior do que no Pré-pago.

Há valores discrepantes superiores nos dois grupos.

Figura 149 - 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.

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: posicionar o cursor em alguma célula do banco de dados, menu Inserir, opção Tabela Dinâmica. O Excel 2007 automaticamente selecionará todas as colunas adjacentes à posição do cursor: no nosso caso, na Plan1, células A1 a N5001 do arquivo Celular01.xlsx, como mostrado na Figura 38. Mas agora queremos pôr os resultados na mesma planilha onde estão os dados, como pode ser visto na Figura 150.

¹⁶ 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.

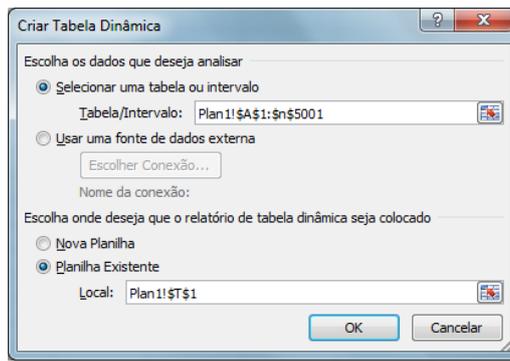


Figura 150 – Seleção de dados da tabela dinâmica – Renda por Tipo

Se pressionarmos OK na Figura 150 o Excel 2007 apresenta a Figura 151, com as várias opções de Layout:

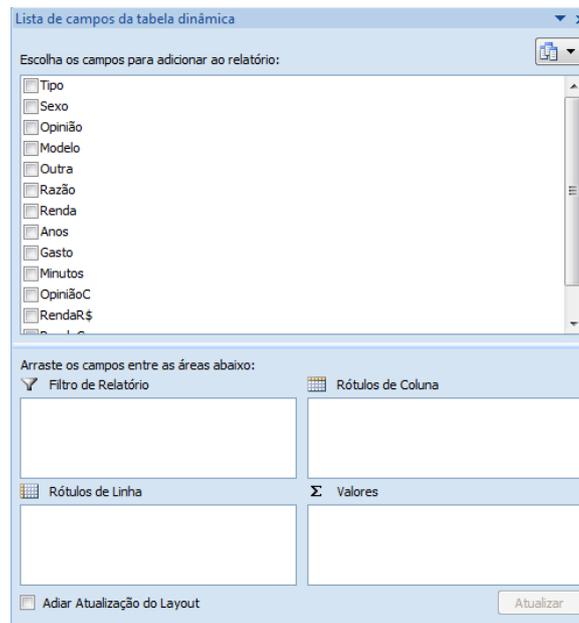


Figura 151 - Layout da tabela dinâmica Renda x Tipo 1a parte

Vamos arrastar a variável Tipo para o campo “Rótulos de Linha”, e Renda para o campo “Σ Valores”, NÃO arraste nenhuma outra variável para o campo “Rótulos de Coluna”, resultando na Figura 152

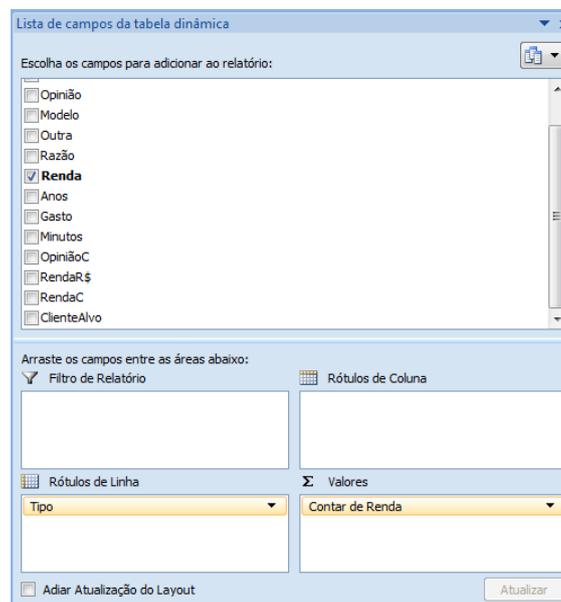


Figura 152 - Layout da tabela dinâmica Renda x Tipo 2a parte

No caso da Figura 152 a tabela apenas contaria quantos clientes há em cada Tipo de serviço. Precisamos arrastar a variável Renda mais vezes até o campo “ Σ Valores” e modificar a forma de resumir os resultados: basta selecionar o botão da variável, e escolher “Configurações do campo de Valor”, ver Figura 153.

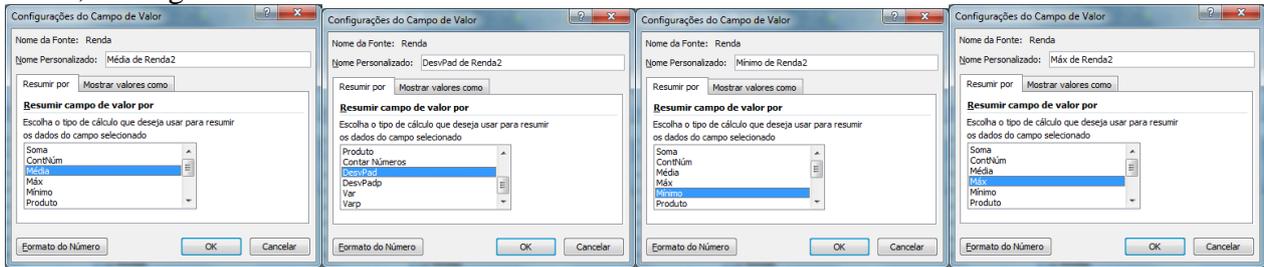


Figura 153 – Formas de resumir os dados de Renda

Usamos “DesvPad” porque os dados do arquivo Celular01.xls referem-se a uma amostra. O layout resultante é mostrado na Figura 154 – Tela à esquerda. E tal como mostrado na Figura 52, precisamos mover o botão Σ Valores do campo Rótulos de Coluna para o campo Rótulos de Linha, resultando na tela à direita da Figura 154.

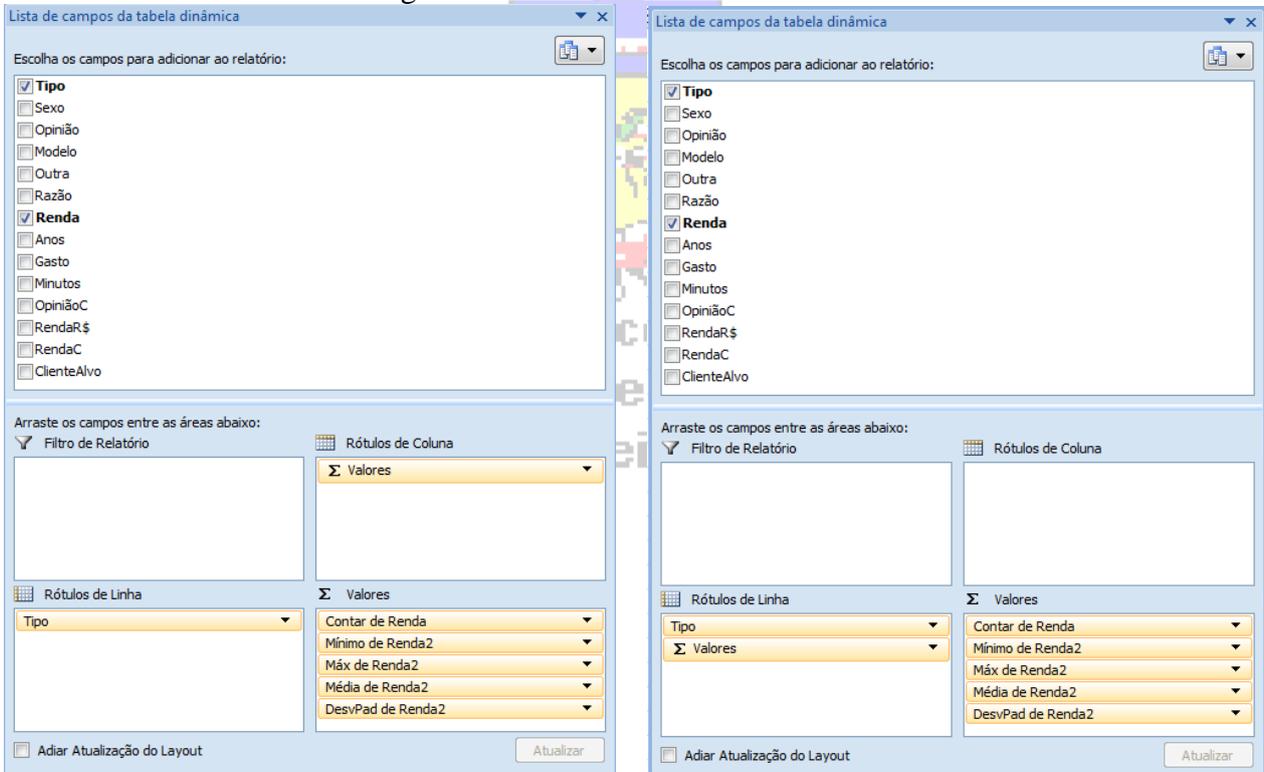


Figura 154 - Layout da tabela dinâmica Renda x Tipo 3a parte

Clicando sobre o X no canto superior direito da tela à direita da Figura 154 chegamos à Figura 155, já sem as células vazias e com modificações para tornar a tabela mais apresentável.

	T	U
1	Tipo de serviço	
2	Pós-pago	
3	Frequências	2102
4	Mínimo de Renda	1,032
5	Máximo de Renda	43,878
6	Média de Renda	8,383899952
7	Desvio padrão de Renda	5,751495488
8	Pré-pago	
9	Frequências	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 Frequências	4993
15	Total Mínimo de Renda	1,005
16	Total Máximo de Renda	35,796
17	Total Média de Renda	5,79693185
18	Total Desvio padrão de Renda	4,992472753

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;

Figura 155 - Tabela dinâmica de Renda por Tipo

- 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 “Filtro de Relatório” na Figura 154, 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 “Filtro de Relatório”, a não ser a crescente dificuldade em interpretar os resultados encontrados. Vamos ver um exemplo, usando a variável Sexo no campo “Filtro de Relatório”, o layout da tabela seria:

Lista de campos da tabela dinâmica

Escolha os campos para adicionar ao relatório:

- Tipo
- Sexo
- Opinião
- Modelo
- Outra
- Razão
- Renda
- Anos
- Gasto
- Minutos
- OpiniãoC
- RendaR\$
- RendaC
- ClienteAlvo

Arraste os campos entre as áreas abaixo:

Filtro de Relatório: Sexo

Rótulos de Coluna: []

Rótulos de Linha: Tipo

Valores: Σ Valores

Valores: Frequências, Mínimo de Renda, Máximo de Renda, Média de Renda, Desvio padrão de Renda

Adiar Atualização do Layout

Atualizar

Figura 156 - Layout da tabela dinâmica Renda x Tipo x Sexo

O resultado pode ser visto na Figura 157:

T		U
1	Sexo	(Tudo)
2		
3	Tipo de serviço	
4	Pós-pago	
5	Freqüências	2102
6	Mínimo de Renda	1,032
7	Máximo de Renda	43,878
8	Média de Renda	8,383899952
9	Desvio padrão de Renda	5,751495488
10	Pré-pago	
11	Freqüências	2891
12	Mínimo de Renda	1,003
13	Máximo de Renda	35,796
14	Média de Renda	3,918023183
15	Desvio padrão de Renda	3,258436647
16	Total Freqüências	4993
17	Total Mínimo de Renda	1,005
18	Total Máximo de Renda	35,796
19	Total Média de Renda	5,79693185
20	Total Desvio padrão de Renda	4,992472753

Figura 157 - Tabela dinâmica de Renda x Tipo x Sexo (geral)

Na Figura 157 a tabela é exatamente igual a da Figura 155, pois não acionamos o filtro da variável Sexo. Na Figura 158 são apresentadas as duas tabelas de Renda por Tipo para o sexo Feminino e Masculino.

T		U	T		U
1	Sexo	Feminino	1	Sexo	Masculino
2			2		
3	Tipo de serviço		3	Tipo de serviço	
4	Pós-pago		4	Pós-pago	
5	Freqüências	709	5	Freqüências	1388
6	Mínimo de Renda	1,032	6	Mínimo de Renda	1,074
7	Máximo de Renda	26,202	7	Máximo de Renda	43,878
8	Média de Renda	6,01730791	8	Média de Renda	9,590888889
9	Desvio padrão de Renda	4,142805999	9	Desvio padrão de Renda	6,078290176
10	Pré-pago		10	Pré-pago	
11	Freqüências	1537	11	Freqüências	1352
12	Mínimo de Renda	1,003	12	Mínimo de Renda	1,005
13	Máximo de Renda	14,592	13	Máximo de Renda	35,796
14	Média de Renda	2,906101563	14	Média de Renda	5,070190828
15	Desvio padrão de Renda	2,110273099	15	Desvio padrão de Renda	3,891982437
16	Total Freqüências	2246	16	Total Freqüências	2740
17	Total Mínimo de Renda	1,003	17	Total Mínimo de Renda	1,005
18	Total Máximo de Renda	13,635	18	Total Máximo de Renda	35,796
19	Total Média de Renda	3,887712121	19	Total Média de Renda	7,358608473
20	Total Desvio padrão de Renda	3,247881088	20	Total Desvio padrão de Renda	5,59309405

Figura 158 - Tabelas dinâmicas de Renda x Tipo x Sexo (Feminino e Masculino)

Podemos calcular os coeficientes de variação percentual e resumi-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 máximos¹⁷ em todas as subdivisões, *presume-se* a existência de valores discrepantes superiores.



¹⁷ 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)