

# ANÁLISE DESCRITIVA E EXPLORATÓRIA DE DADOS USANDO O EXCEL

## Algumas dicas

Pedro Alberto Barbeta

Estas notas foram escritas baseadas no Excel versão de 2002.

### 1 - CRIAÇÃO DO ARQUIVO DE DADOS

Ao abrir o *Excel*, aparece uma planilha (uma matriz de linhas e colunas). Cada célula é identificada pela letra da coluna e número da linha correspondente. Por exemplo, **A1** refere-se à célula da *coluna A* e *linha 1* (canto superior esquerdo da planilha).

Sugere-se que se use a primeira linha para colocar os nomes das variáveis. Assim, as colunas indicam as variáveis e as linhas os registros dos casos observados. É usual identificar os possíveis resultados de cada variável com códigos numéricos (veja como exemplo o arquivo *SacoGrande*, planilha *Dados*):

B	C	D	E	F
local	p.a.p	instr.	tam.	renda
1	0	3	4	10,3
1	0	3	4	15,4
1	1	2	4	9,6
1	0	2	5	5,5
1	1	3	4	9
1	1	1	1	2,4
1	0	3	2	4,1

É conveniente formatar os códigos de variáveis qualitativas como texto. Para isso, entrar em *Formatar* (no Menu principal, topo da tela), *Célula*, *Texto*. Alternativamente, pode-se usar a função SE, como é descrito a seguir.

### FUNÇÃO "SE" (tópico opcional)

Para alterar os códigos numéricos das variáveis qualitativas em rótulos, conforme aparece na planilha *Dados2*, pode-se usar a função SE. Para isso, no Menu (no alto da tela) clicar em *Inserir*, *Função*. Da lista de funções que aparece, clicar na categoria de função *Lógica* (lado esquerdo) e função *SE* (lado direito da caixa de diálogo).

Para facilitar, na planilha *Dados2*, copiamos os dados das variáveis qualitativas nas colunas H, I e J. Para a variável P.A.P. (coluna I), que só tem duas

categorias (1=usa, 2=não usa), só usamos uma vez a função SE. Colocamos o cursor na célula C2 e, na caixa de diálogo da função SE, preenchemos os seguintes parâmetros:

Teste_lógico:	I2=1
Valor_se_verdadeiro:	usa
Valor_se_falso:	não usa

Observamos que I2 é o local onde está o primeiro código numérico de P.A.P. Observar que ao digitar os parâmetros da função SE, acima da planilha aparece a fórmula analítica dessa função:

=SE(I2=1;"usa";"não usa")

Esta expressão também aparece se você clicar sobre a célula C2 (onde digitamos esta expressão). Para deixar a função, clicar *Esc*.

Para estender a aplicação dessa função nas outras células da coluna C2, basta colocar o cursor no canto direito de baixo da célula em que se digitou a expressão e arrastar com o *mouse* para baixo (com o botão do lado esquerdo apertado).

Para a variável ***instr.***, que tem mais de duas categorias, torna-se necessário aninhar comandos SE. Com o cursor sobre a célula D2, digitou-se o seguinte comando:

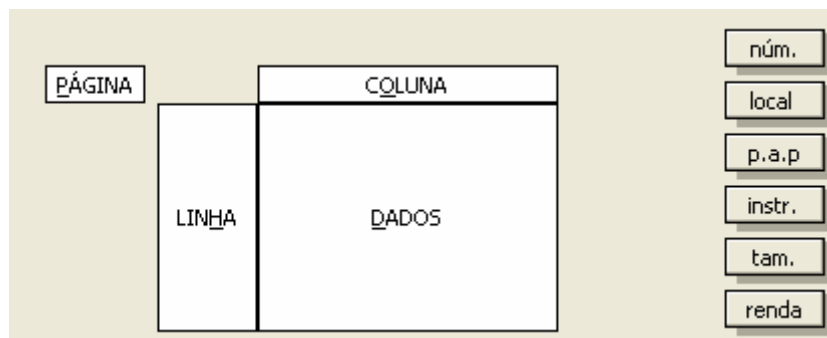
=SE(J2=1;"nenhum";SE(J2=2;"primeiro grau";"segundo grau"))

E depois, arrastou-se o *mouse* para as demais células da coluna D.

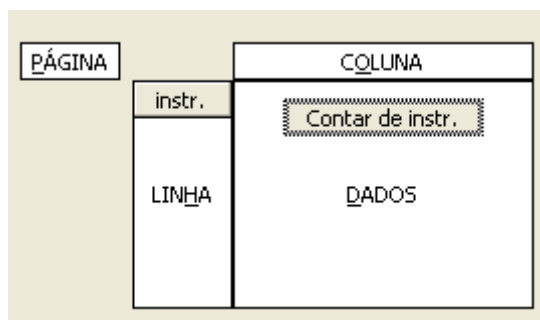
## **2 - DISTRIBUIÇÃO DE FREQUÊNCIAS DE VARIÁVEIS QUALITATIVAS – CLASSIFICAÇÃO SIMPLES**

Com o cursor sobre a tabela, clicar em *Dados* (menu principal) e *Relatório da Tabela Dinâmica e Avançar*. Aparece uma janela para que apontemos o intervalo de entrada. Se o Excel estiver apontando para toda a tabela, ótimo. Se não, arrastar o cursor sobre as variáveis que você pretende analisar (não tem problema se incluir mais variáveis).

Clicar em *Avançar*. Aparece uma tela com uma tabela a ser montada e os nomes das variáveis escolhidas do lado direito. Ver figura a seguir (se não aparecer essa figura, clicar em *Layout*).



Escolha a variável que você pretende analisar e arraste-a até a margem esquerda (Linha) da tabela que está sendo montada. Arraste-a novamente para o centro da tabela. Nesta posição deverá aparecer *Contagem de...*. Caso contrário, clique duas vezes nesta posição e escolha *Cont.Núm.* Em seguida, clique em *Avançar* e *Concluir* (ver exemplo na figura a seguir). A tabela será apresentada em nova planilha.



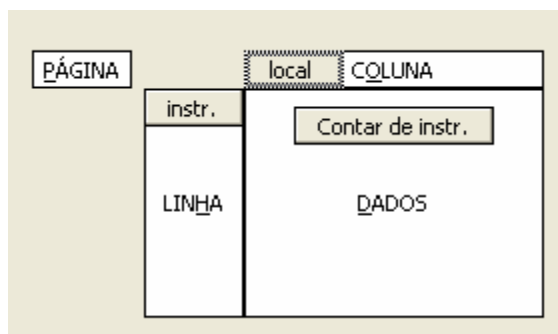
Se você quiser incluir frequências relativas (porcentagens), copie a tabela para outra posição e, com o cursor sobre ela, clique em *Dados, Relatório da Tabela Dinâmica, Layout*. Onde está *Contagem de...*, clique duas vezes. Clique em *Opções*. Onde está *Normal*, substitua por *% da coluna*. Clique em *OK, Avançar* e *Concluir*.

Para construir um gráfico adequado, deixe o cursor numa célula fora da tabela e aperte no desenho de um gráfico na caixa de ferramentas (alternativamente, *Inserir, Objeto, Gráfico...*). Escolha em tipos de gráficos, Colunas, Barras ou Pizza. Ao ser solicitado sobre o intervalo de entrada, escolha as células da tabela que contém as informações do gráfico. Daí, preencher adequadamente as diversas opções que as caixas de diálogo fornecem.

**NOTA.** Normalmente o gráfico ou tabela fornecida pelo Excel não é exatamente aquele que precisamos ou que queremos – é necessário ajustá-lo.

### 3 - DISTRIBUIÇÃO DE FREQUÊNCIAS DE VARIÁVEIS QUALITATIVAS – CLASSIFICAÇÃO DUPLA

Para a construção de distribuições de frequências conjunta (tabelas de contingência), seguir os mesmos passos do caso anterior, mas preencher também a Coluna na tabela que está sendo montada. O Layout a seguir gera uma tabela de dupla entrada (tabela de contingência) entre Local e Grau de Instrução.



### 4 - DISTRIBUIÇÃO DE FREQUÊNCIAS DE VARIÁVEIS DISCRETAS

Uma tabela de distribuição de frequências de uma variável discreta pode ser feita de forma análoga à distribuição de frequências de variáveis qualitativas (*Dados, Relatório de tabela e gráficos dinâmicos*).

### 5 - DISTRIBUIÇÃO DE FREQUÊNCIAS DE VARIÁVEIS CONTÍNUAS

O Excel não é uma boa ferramenta para fazer um histograma de frequências, mas vamos lá...

Faremos uso da ferramenta *Análise de Dados*. Se, ao clicar *Ferramenta* (no menu principal), não aparecer *Análise de Dados*, clique em *Suplementos* e inclua o módulo *Ferramentas de Análise*. Da próxima vez em que clicar em *Ferramentas* aparecerá *Análise de Dados*.

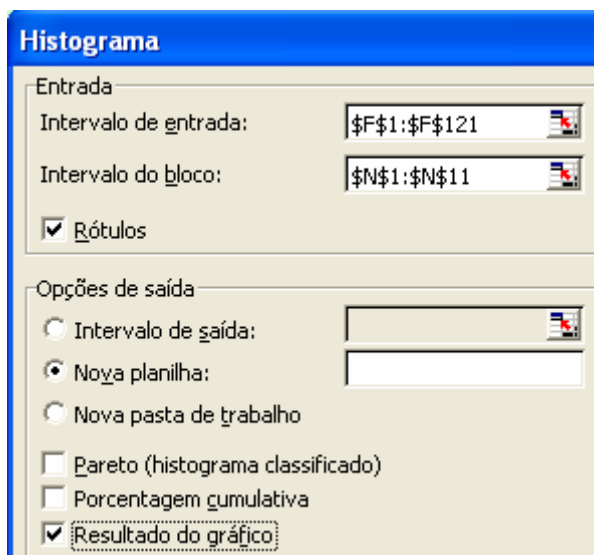
É recomendável especificar as classes. Para se ter uma idéia da escolha das classes, pode-se ordenar os dados com respeito à variável em análise. Para isto, deixe o cursor sobre uma célula da coluna em que estão os dados da variável de interesse. Clique em *Dados* (menu principal) e *Classificar*. Na caixa de diálogo, em *Classificar por*, você pode escolher a variável que deseja ordenar. Com os dados ordenados, escolher uma divisão adequada para as classes. Por exemplo, a variável

*Renda* tem seu menor valor 0,1 e maior valor 25,7. Se tomarmos classes 0 — 3 , 3 — 6, ..., então teremos 9 classes (um número razoável!).

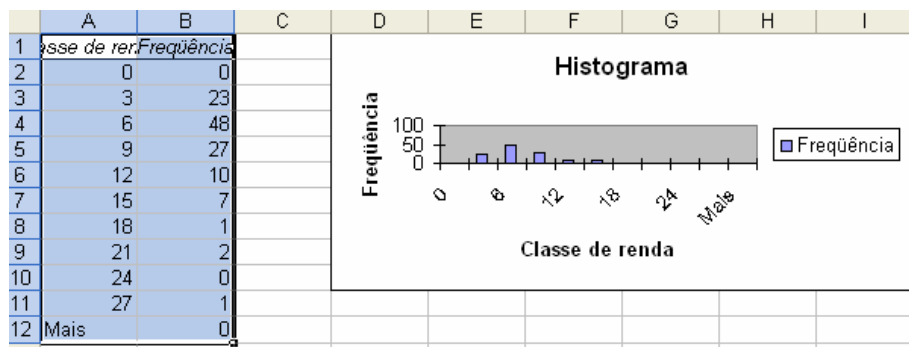
Coloca-se, então, a lista dos limites inferiores das classes em uma coluna (pode ser na mesma planilha ou numa planilha nova). Colocamos a seguinte seqüência na coluna N:

N
classes de renda
0
3
6
...
27

Vá ao menu principal e clique em *Ferramentas, Análise de Dados*. O *Excel* abre uma janela com várias ferramentas estatísticas. Escolha *Histograma*. Será aberta uma caixa de diálogo pedindo o *Intervalo de Entrada*. Arraste o cursor sobre os dados da variável em análise (pode incluir o rótulo). Em *Intervalo do Bloco*, arraste o cursor sobre a lista de classes que você criou (se incluiu o rótulo no intervalo de entrada, você deve incluí-lo aqui também e, neste caso, marcar a janela *Rótulos*). Em opções de saída, marcar *Resultado do gráfico* (ver figura a seguir).



O *Excel* abre uma nova planilha com a tabela e o gráfico de freqüências. Ver figura a seguir.



Note que está horrível! Vários ajustes são necessários.

Na tabela, excluir a linha com a categoria "*mais*". Observa-se que a frequência colocada ao lado das classes representa a contagem do número de observações entre o número da coluna imediatamente anterior até o valor da coluna atual. Por exemplo, na frente do "3" estará a frequência da classe "0 — 3", na frente do "6", a frequência da classe "3 — 6", e assim por diante.

No gráfico, é necessário expandi-lo, arrastando o cursor sobre as suas bordas. Clicar sobre a legenda (*freqüência*) e deletá-la. Para juntar as colunas, como num histograma, basta clicar 2 vezes sobre elas. O Excel abre uma caixa de diálogo chamada "Formatar seqüência de dados". Clicar 2 vezes em *Opções*. Em *Largura do Espaçamento*, colocar 0 (zero).

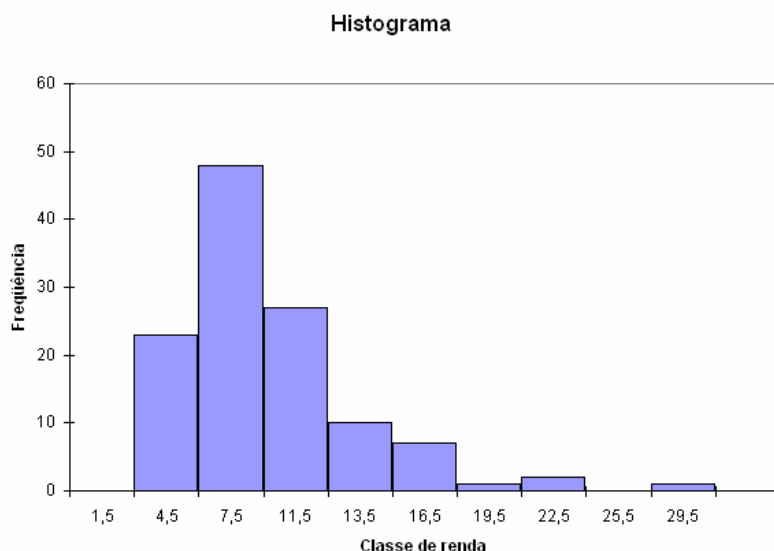
Há um certo deslocamento na escala horizontal do gráfico. O *Excel* indica o limite superior da classe no meio das colunas (o que está errado). Pode-se corrigir isto, substituindo os valores limites das classes pelos respectivos pontos médios. No exemplo da renda, substituir:

0      3      6      9      ...

por

1,5    4,5    7,5    10,5    ...

Podem-se também adequar os títulos dos eixos simplesmente clicando em cima desses. Da mesma forma, podem-se adequar as fontes, que normalmente aparecem muito grandes. Veja o gráfico após os ajustes.



## 6 – MEDIDAS DESCRITIVAS

Há duas maneiras de se obter medidas descritivas (média, desvio padrão, etc.) pelo Excel. Se você deseja apenas uma medida, pode-se colocar o cursor numa célula vazia da planilha e clicar em  $f_x$  (ou, no Menu principal, clicar em *Inserir, Função*). Nas funções estatísticas, escolher a função desejada (p. ex., MÉDIA, DESVPAD, QUARTIL, etc.).

Quando você pretende ter várias medidas descritivas dos dados de uma variável, pode clicar em *Ferramentas, Análise de Dados e Estatística Descritiva*. Na caixa de diálogo, preencher o Intervalo de Entrada arrastando o mouse sobre os dados que deseja analisar. Nas opções de saída, marcar *Resumo Estatístico*. O *Excel* abre uma nova planilha com várias medidas descritivas.

## 7 – MEDIDAS DESCRITIVAS EM SUBGRUPOS DA AMOSTRA

Muitas vezes, tem-se o interesse em obter medidas descritivas de uma variável quantitativa em subgrupos formados por categorias de uma ou mais variáveis qualitativas. Por exemplo, medidas descritivas de *Renda* por *Localidade*.

Neste caso, pode-se usar Relatório da Tabela Dinâmica como discutido no tópico 2, onde as variáveis qualitativas são colocadas na *Linha* e/ou *Coluna* da tabela e a variável quantitativa no centro (*Dados*). No lugar de *contagem de...*, escolher a estatística de interesse. Para mais de uma medida descritiva, arrastar novamente a variável quantitativa no centro (*Dados*) e escolher a outra medida.

Por exemplo, no layout abaixo o Excel vai criar uma tabela com a média e o desvio padrão de Renda para cada Localidade.

