

Aplicações Computacionais em Exploração e Análise de Dados: Introdução ao Excel

Lupércio F. Bessegato & Camila B. Zeller & Marcel T. Vieira

UFJF – Departamento de Estatística
2013



Apresentação

- Lupércio França Bessegato
lupercio.bessegato@ufjf.edu.br
Departamento de Estatística
- Camila Borelli Zeller
camila.zeller@ufjf.edu.br
Departamento de Estatística
- Marcel de Toledo Vieira
marcel.vieira@ufjf.edu.br
Departamento de Estatística

Roteiro

1. Introdução
2. Tabelas de Frequência
3. Apresentação Gráfica
4. Medidas-resumo
5. Análise Exploratória Univariada

Planilha Excel®

Representação em Tabela

Dia	Paulo	Joana	Patrícia	Pedro
1	8,5	8,2	7,9	7,4
2	7,8	8,1	7,5	8,9
3	8,3	8,4	9	8,4
4	8	6,9	9,6	7,6
28	7,5	8,4	7,9	8,5
29	6	7,6	8,6	8,1
30	8,6	8,7	8,3	8,6
31	8,9	8,1	8,5	8,4
Total	63,6	64,4	67,3	65,9

Exemplo de Planilha

√ Planilha: tutorial.xls/guia

Area de Trabalho | Fonte | Abntamento

E14 | fx | =SOMA(E5:E12)

	A	B	C	D	E	F
1	Relação de Livros					
2						
3	Título	Autor	Editora	Tipo	Preço de compra	
4						
5	A moreninha	José de Alencar	Grandes Publicações	Romance	21,35	
6	Agosto	Rubem Fonseca	Política Nacional	Ficção	45,1	
7	As Valkírias	Paulo Coelho	Rocco	Ficção	45,15	
8	Dom Casmurro	Machado de Assis	Grandes Publicações	Romance	29,65	
9	Estação Carandiru	Dráuzio Varella	Verdade	Não ficção	44,25	
10	Gabriela	Jorge Amado	Grandes Publicações	Romance	42,99	
11	Mar sem fim	Amr Klink	Navegar é Preciso	Não ficção	35	
12	O alquimista	Paulo Coelho	Rocco	Ficção	39,4	
13						
14				Total	302,89	
15						
16						
17						

Banco de Dados

- Guia com o conjunto de dados deve ser preservada!
- Nomear o banco de dados para uso posterior
 - √ Selecionar o conjunto de dados
 - √ Clicar com botão direito: "Definir nome"
 - √ Nome: **banco**

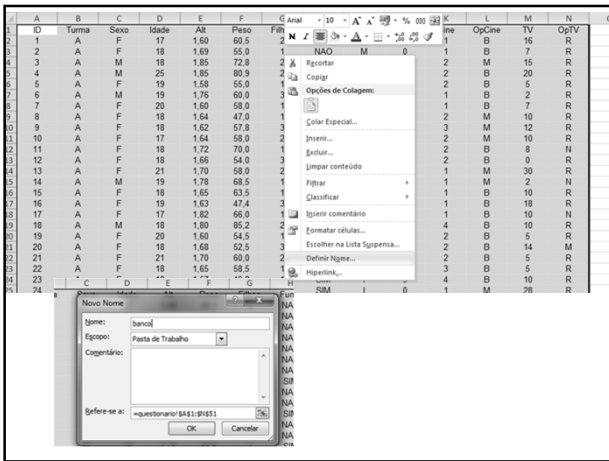
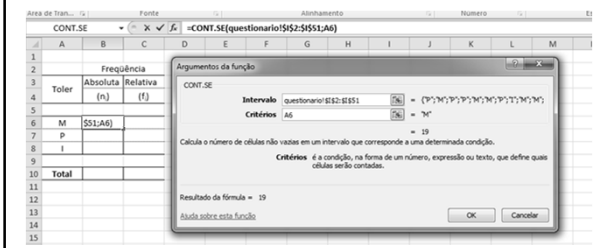


Tabela de Frequência – Variável Categórica

- Contagem de cada nível da variável Toler
- Comando **CONT. SE** (usar Inserir função)



- Gráficos

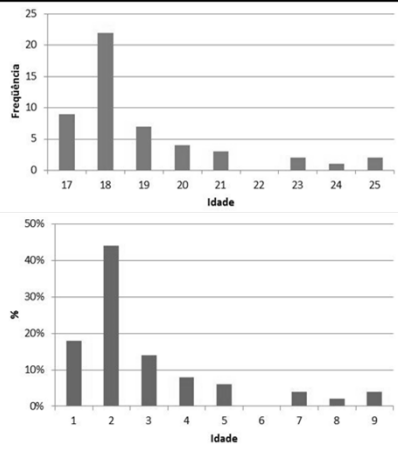


Tabela de Frequência – Variável Contínua

- Contagem para cada valor da variável Peso
 - ✓ Usar comando **FREQUÊNCIA** (usar Inserir função)

Argumentos da função

FREQUÊNCIA

Matriz_dados: questionario!F2:F51 = {80; 5; 55; 72; 8; 80; 5; 55; 60; 58; 47; 57; ...}

Matriz_bin: C4:C8-0,0000000001 = {49,99999999999999; 59,99999999999999; 69,99999999999999; 79,99999999999999; 89,99999999999999}

Calcula a frequência de ocorrência de valores em um intervalo de valores e retorna uma matriz vertical de números contendo um elemento a mais do que 'Matriz_bin'.

Matriz_bin é uma matriz de ou referência a intervalos nos quais você deseja agrupar os valores contidos em 'Matriz_dados'.

Importante: Concluir com Ctrl+Shift+Enter

- Construir tabela

	A	B	C	D	E	F	G
1							
2		Peso	n_i	f_i	f_{cc}		
3							
4	40	50	8	16,0%	16,0%		
5	50	60	22	44,0%	60,0%		
6	60	70	8	16,0%	76,0%		
7	70	80	6	12,0%	88,0%		
8	80	90	5	10,0%	98,0%		
9	90	100	1	2,0%	100,0%		
10							
11	Total		50		100,0%		
12							

- Construir gráfico

- ✓ Selecionar valores tabela
- ✓ Colunas

Arquivo Página Inicial Inserir Layout da Página Fórmulas Dados Revisão Exibição

Coluna 2D

Coluna 3D

Cilindro

Cone

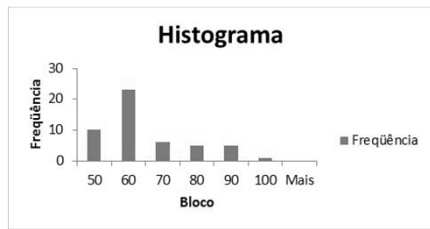
Prisma

Índices de Tipos de Gráfico...

• Dados > Análise de Dados > Histograma

	Peso	n	f	f _{ac}
4	40 - 50	8	16,0%	16,0%
5	50 - 60	22	44,0%	60,0%
6	60 - 70	8	16,0%	76,0%
7	70 - 80	6	12,0%	88,0%
8	80 - 90	5	10,0%	98,0%
9	90 - 100	1	2,0%	100,0%
	Total	50	100,0%	

• Gráfico:



Bancos de Dados

Funções de Banco de Dados

- BDMÉDIA
- BDESVPA (ou BDDEST)
- BDMÁX
- BDMÍN
- BDCONTAR
- BDSOMA

Funcionamento Funções de Banco de Dados

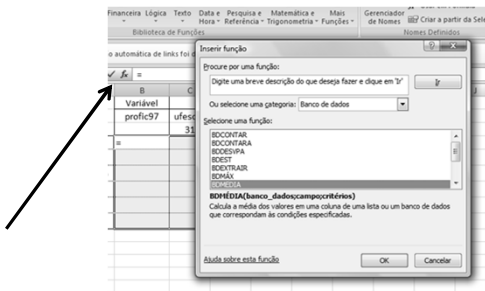
- Preparar uma tabela para cálculo de medidas:

	A	B	C	D
1		Variável	Critério	
2		profic97	ufesc_c	ufesc_c
3			31	35
4	Média			
5	Desvio-padrão			
6	Coefficiente de variação			
7	Mínimo			
8	Máximo			
9	Quantidade (n)			
10				

- Acesso às funções Banco de Dados:

√ Serão usadas as funções:

BDMÉDIA; BDESVPA; BDMÍN; BDMÁX; BDCONTARA



- Use o tutorial do Excel para montar as fórmulas!

Análise Exploratória de Dados

O que é Análise Exploratória de Dados?

- Uma filosofia/abordagem para análise de dados
- Emprega uma variedade de técnicas (a maioria gráficas)...Neste curso, trabalhamos com alguns deles:
 - √ Diagrama de dispersão
 - √ **Ramo e folhas (p/ conhecer)**
 - √ **Boxplot**
 - √ Individual Plot

Técnicas que buscam:

- maximizar o “insight” do conjunto de dados;
- perceber a estrutura subjacente;
- extrair variáveis importantes;
- detectar valores atípicos (extremos) e anomalias;
- testar hipóteses fundamentais;
- desenvolver modelos parcimoniosos; e
- determinar conjunto ótimo de fatores

ideia Básica

- Modelo = Suave + Irregular (tosco)
- Técnicas visuais podem frequentemente separar mais o “suave” do “irregular” (“ruído”)

Clássica vs. Exploratória

- Sequencia Clássica:
 - √ Problema > Dados > Modelo > Análise > Conclusões
- Exploratória:
 - √ Problema > Dados > Análise > Modelo > Conclusões

Tratamento de Dados

- Clássica:
 - √ Média e desvio padrão = estimativas pontuais
 - √ Medida de variabilidade explicada – r de Pearson
- Exploratória
 - √ Resumo Numérico (5): Min, Q1, Median, Q3, Max
 - √ todos (maioria) dados=resumos visuais
 - √ Dispersão
 - √ Histograma
 - √ Boxplot

Análise Descritiva

- Inicia-se quase sempre pela verificação dos tipos disponíveis de variáveis
- Elas podem ser resumidas por tabelas, gráficos e/ou medidas

Objetivos

- Familiarização com os dados
- Detecção de estruturas interessantes
- Presença de valores atípicos (*outliers*)

- Todos estes aspectos foram tratados neste curso!

Referências

Bibliografia

- Magalhães, M.N. e Lima, A.C.P.L. (Edusp)
Noções de Probabilidade e Estatística
- Wild, C.J. e Seber, G.A.F. (LTC)
Encontros com o Acaso: um Primeiro Curso de Análise de Dados e Inferência
- Agresti, A. e Agresti, B.F. (Dellen Pub.)
Statistical Methods for the Social Sciences
