Introdução à Análise de Dados I

Lupércio F. Bessegato & Marcel T. Vieira

UFJF – Departamento de Estatística
2013

Apresentação

- Lupércio França Bessegato
  lupercio.bessegato@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

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!

Estudo de Caso – SAEB

Estudo de Caso - SAEB

- Dados do Sistema Nacional de Avaliação da Educação Básica – SAEB
  - √ Ano de 1999
  - √ Variáveis incluem:
    - Escores de proficiência
    - Condições sócio-econômicas-demográficas
- Objetivo:
  - √ Manipulação, análise e interpretação de dados
- Banco de dados: *Saeb99_Mestrado.xlsx/banco*

- Variáveis e suas medidas

ufesc_c	*Estado onde se localiza a Escola*	*31 - Minas Gerais* *35 - São Paulo*
codalu_c	*Código do Aluno*	
q12_3	*Que idade você tinha quando entrou na 1a série?*	*1 - Menos de 5 anos* *2 - 5 anos* *3 - 6 anos* *4 - 7 anos* *5 - 8 anos* *6 - 9 anos ou mais*
q13_3	*Qual o seu sexo?*	*1 - Masculino* *2 - Feminino*
q14_3	*Como você se considera?*	*1 - branco(a)* *2 - pardo(a)/mulato(a)* *3 - negro(a)* *4 - amarelo(a)* *5 - indígena*
q17_3	*Qual o nivel de instrução do seu pai?*	*1 - Nunca freqüentou a escola* *2 - Ensino Fundamental (1o Grau) - 1a à 4a série* *3 - Ensino Fundamental (1o Grau) - 5a à 8a série* *4 - Ensino Médio (2o Grau)* *5 - Superior* *6 - Pós-Graduação* *7 - Não Sei*

q18_3	*Qual o nivel de instrução de sua mãe?*	*(mesma codificação da variável anterior)*
q111_3	*Como é a sua situação econômica?*	*1 - Eu não trabalho e dependo economicamente da minha familia* *2 - Eu trabalho, mas dependo economicamente da minha familia* *3 - Eu trabalho e não dependo da minha familia para me manter* *4 - Eu trabalho e outras pessoas dependem de meu trabalho para viver*
q114_3	*Você já deixou de freqüentar a escola por algum período?*	*1 - Não* *2 - Sim, por 1 ano* *3 - Sim, por 2 anos* *4 - Sim, por 3 anos* *5 - Sim, por 4 anos* *6 - Sim, por mais de 4 anos*
q116_3	*Você já repetiu de ano?*	*1 - Não* *2 - Sim, 1 vez* *3 - Sim, 2 vezes* *4 - Sim, 3 vezes* *5 - Sim, 4 vezes* *6 - Sim, mais de 4 vezes*
q121_3	*Quanto tempo você leva para chegar de sua casa até a escola?*	*1 - Menos de 15 minutos* *2 - De 15 a 30 minutos* *3 - De 31 minutos a 1 hora* *4 - Mais de 1 hora*
q26_3	*Você gosta de Fisica?*	*1 - Não gosto* *2 - Gosto mais ou menos* *3 - Gosto muito*
profic99	*Proficiência do Aluno em Fisica*	

Análise Exploratória – Objetivos

√ Análise e comparação das variáveis relevantes da base de dados
(resumo por meio de gráficos, tabelas e medida-resumo)

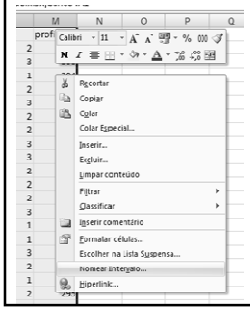
√ Identificação de estratificações importantes ao estudo de variáveis relevantes do banco de dados;

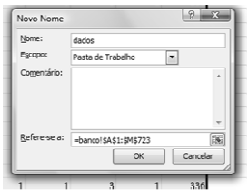
√ Identificação de padrões de variáveis relevantes do banco de dados;

√ Estabelecimento de hipóteses para estudo posterior mais detalhado

- 1º Passo:
  - √ Abrir a planilha *Saeb99_Mestrado.xlsx*
  - √ Copiar o banco de dados para uma nova planilha
  - √ Nomear a guia com o banco com um nome especial (*banco*, por exemplo)
  - √ Salvar a nova planilha para que você possa trabalhar nela, preservando os dados originais

- 2º Passo:
  - √ Selecione o banco em sua planilha de trabalho
  - √ Nomeie o intervalo selecionado (denomine-o, por exemplo, *dados*). Use o botão direito do mouse





Item 1 – Desempenho por Estado

√ Análise da proficiência de todos os aluno:
- Tendência central (média)
- Dispersão (desvio padrão)
- Assimetria (histograma)

√ Análise da proficiência por estado
- Repetir o procedimento anterior

√ Pergunta importante durante o exercício:
- Há diferença na proficiência por estado?

Bancos de Dados

Funções de Banco de Dados

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

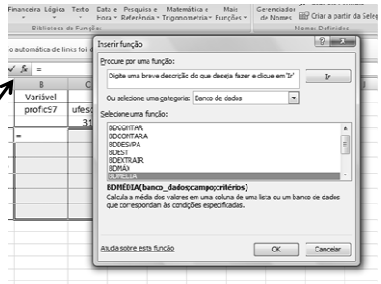
Funcionamento Funções de Banco de Dados

- 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	Coeficiente de variação			
7	Mínimo			
8	Máximo			
9	Quantidade (n)			
10				

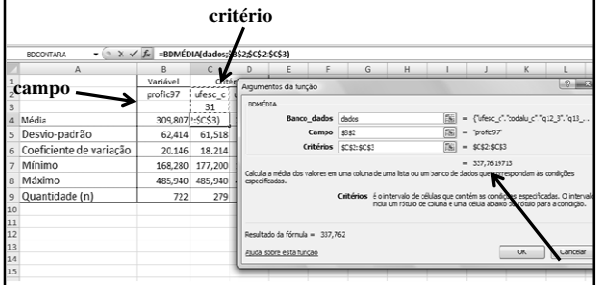
- Acesso à funções Banco de Dados:
  - √ Serão usadas as funções:

    BDMÉDIA; BDDESVPA; BDMÍN; BDMÁX; BDCONTARA

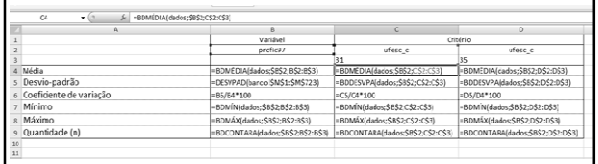


- Use o tutorial do Excel para montar as fórmulas!

- Cálculo da Média por Estado:



- Procedimento similar para calcular as próximas medidas
  - √ Montar as fórmulas de uma coluna (C, por exemplo) e copiar para outra (D).
- Fórmulas da montagem da tabela



	A	B	C	D
1		Variável	Critério	
2		profic97	ufesc_c	ufesc_c
3			31	35
4	Média	=BDMEDIA(dados;$B$2:B$2;B$3)	=BDMEDIA(dados;$B$2:C$2;C$3)	=BDMEDIA(cados;$B$2:D$2;D$3)
5	Desvio-padrão	=DESVPAD(barco $M$1:$M$723)	=BDDESVPA(dados;$B$2:C$2;C$3)	=BDDESVPA(dados;$B$2:D$2;D$3)
6	Coeficiente de variação	=B5/B4*100	=C5/C4*100	=D5/D4*100
7	Mínimo	=BDMÍN(dados;$B$2:B$2;B$3)	=BDMÍN(dados;$B$2:C$2;C$3)	=BDMÍN(dados;$B$2:D$2;D$3)
8	Máximo	=BDMÁX(dados;$B$2:B$2;B$3)	=BDMÁX(dados;$B$2:C$2;C$3)	=BDMÁX(dados;$B$2:D$2;D$3)
9	Quantidade (n)	=BDCONTARA(dados;$B$2:B$2;B$3)	=BDCONTARA(dados;$B$2:C$2;C$3)	=BDCONTARA(dados;$B$2:D$2;D$3)
10				
11				

- Resultados:

	Variável	Critério	
	profic97	ufesc_c	ufesc_c
		31	35
Média	309,807	327,762	292,201
Desvio-padrão	62,414	61,518	56,166
Coeficiente de variação	20,146	18,214	19,222
Mínimo	168,280	177,200	168,280
Máximo	485,940	485,940	475,910
Quantidade (n)	722	279	443

- Há diferença na proficiência entre os estados? É significativa?
  - √ Em qual estado o desempenho é melhor?
  - √ Onde há maior homogeneidade?

Item 1 – Histograma

- Montagem da tabela de frequência:
  - √ (Máximo; Mínimo) = (168,28; 485,94)
  - √ Intervalo de classes: 10
  - √ Limite de classe inferior (1º): 165
  - √ Limite de classe superior (último): 495
  - √ Quantidade de intervalos: 33
- Tabelas de frequência:
  - √ *case_saeb_sol.xlsx/tabela_a_estado*

- Uso da Tabela
  - √ Especificar na Tabela
    - variável contínua em análise (A3): *profic99*
    - variável de estratificação (D2): *ufesc_c*
    - Níveis da variável categórica (D3 e F3): 31 e 35

profic99			Frequências ufesc_c 31 n_i	31 f_i	35 n_i	35 f_i	Total n_i	Total f_i	Frequências Acumuladas ufesc_c 31 n_{ac}	31 f_{ac}	35 n_{ac}	35 f_{ac}	Total n_{ac}	Total f_{ac}
165	├─	175	0	0,0%	1	0,2%	1	0,1%	0	0,0%	1	0,2%	1	0,1%
175	├─	185	2	0,7%	2	0,5%	4	0,6%	2	0,7%	3	0,7%	5	0,7%
185	├─	195	2	0,7%	3	0,7%	5	0,7%	4	1,4%	6	1,4%	10	1,4%
195	├─	205	2	0,7%	7	1,6%	9	1,2%	6	2,2%	13	2,9%	19	2,6%
205	├─	215	4	1,4%	12	2,7%	16	2,2%	10	3,6%	25	5,6%	35	4,8%
215	├─	225	5	1,8%	31	7,0%	36	5,0%	15	5,4%	56	12,6%	71	9,8%
225	├─	235	4	1,4%	20	4,5%	24	3,3%	19	6,8%	76	17,2%	95	13,2%
235	├─	245	7	2,5%	32	7,2%	39	5,4%	26	9,3%	108	24,4%	134	18,6%
245	├─	255	4	1,4%	29	6,5%	33	4,6%	30	10,8%	137	30,9%	167	23,1%
255	├─	265	10	3,6%	23	5,2%	33	4,6%	40	14,3%	160	36,1%	200	27,7%

- √ Especificar nos critérios para estratificação e contagem:
  - 1º Limite inferior (C45): 165
  - Intervalo de classe (C44): 10

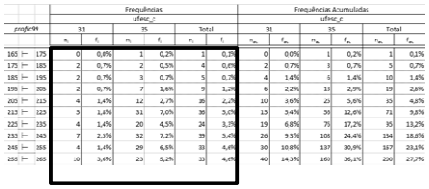
	A	B	C	D	E	F	G	H
41								
42								
43				Criterio 1		Criterio 2		Total
44		int.	10	profic99	ufesc_c	profic99	ufesc_c	profic99
45			175	<175	31	<175	35	<175
46			185	<185	31	<185	35	<185
47			195	<195	31	<195	35	<195
48			205	<205	31	<205	35	<205
49			215	<215	31	<215	35	<215
50			225	<225	31	<225	35	<225
51			235	<235	31	<235	35	<235
52			245	<245	31	<245	35	<245
53			255	<255	31	<255	35	<255
54			265	<265	31	<265	35	<265
55			275	<275	31	<275	35	<275
56			285	<285	31	<285	35	<285
57			295	<295	31	<295	35	<295
58			305	<305	31	<305	35	<305
59			315	<315	31	<315	35	<315

- Tabela de frequências – Proficiência:

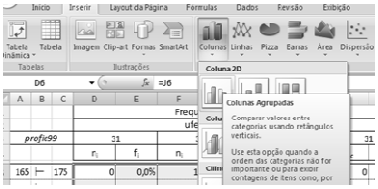
profic99			Frequências ufesc_c 31 n_i	31 f_i	35 n_i	35 f_i	Total n_i	Total f_i	Frequências Acumuladas ufesc_c 31 n_{ac}	31 f_{ac}	35 n_{ac}	35 f_{ac}	Total n_{ac}	Total f_{ac}
165	├─	175	0	0,0%	1	0,2%	1	0,1%	0	0,0%	1	0,2%	1	0,1%
175	├─	185	2	0,7%	2	0,5%	4	0,6%	2	0,7%	3	0,7%	5	0,7%
185	├─	195	2	0,7%	3	0,7%	5	0,7%	4	1,4%	6	1,4%	10	1,4%
195	├─	205	2	0,7%	7	1,6%	9	1,2%	6	2,2%	13	2,9%	19	2,6%
205	├─	215	4	1,4%	12	2,7%	16	2,2%	10	3,6%	25	5,6%	35	4,8%
215	├─	225	5	1,8%	31	7,0%	36	5,0%	15	5,4%	56	12,6%	71	9,8%
225	├─	235	4	1,4%	20	4,5%	24	3,3%	19	6,8%	76	17,2%	95	13,2%
235	├─	245	7	2,5%	32	7,2%	39	5,4%	26	9,3%	108	24,4%	134	18,6%
245	├─	255	4	1,4%	29	6,5%	33	4,6%	30	10,8%	137	30,9%	167	23,1%
255	├─	265	10	3,6%	23	5,2%	33	4,6%	40	14,3%	160	36,1%	200	27,7%
265	├─	275	11	3,9%	26	5,9%	37	5,1%	51	18,3%	186	42,0%	237	32,8%
275	├─	285	6	2,2%	28	6,3%	34	4,7%	57	20,4%	214	48,3%	271	37,5%
285	├─	295	13	4,7%	28	6,3%	41	5,7%	70	25,1%	242	54,6%	312	43,2%
295	├─	305	15	5,4%	27	6,1%	42	5,8%	85	30,5%	269	60,7%	354	49,0%
305	├─	315	7	2,5%	25	5,6%	32	4,4%	92	33,0%	294	66,4%	386	53,5%
315	├─	325	15	5,4%	13	2,9%	28	3,9%	107	38,4%	307	69,3%	414	57,3%
325	├─	335	15	5,4%	26	5,9%	41	5,7%	122	43,7%	333	75,2%	455	63,0%
335	├─	345	21	7,5%	24	5,4%	45	6,2%	143	51,3%	357	80,6%	500	69,3%
345	├─	355	13	4,7%	23	5,2%	36	5,0%	156	55,9%	380	85,8%	536	74,2%
355	├─	365	19	6,8%	19	4,3%	38	5,3%	175	62,7%	399	90,1%	574	79,5%
365	├─	375	15	5,4%	8	1,8%	23	3,2%	190	68,1%	407	91,9%	597	82,7%
375	├─	385	23	8,2%	11	2,5%	34	4,7%	213	76,3%	418	94,4%	631	87,4%
385	├─	395	11	3,9%	7	1,6%	18	2,5%	224	80,3%	425	95,9%	649	89,9%
395	├─	405	18	6,5%	8	1,8%	26	3,6%	242	86,7%	433	97,7%	675	93,5%
405	├─	415	12	4,3%	3	0,7%	15	2,1%	254	91,0%	436	98,4%	690	95,6%
415	├─	425	13	4,7%	2	0,5%	15	2,1%	267	95,7%	438	98,9%	705	97,6%
425	├─	435	6	2,2%	1	0,2%	7	1,0%	273	97,8%	439	99,1%	712	98,6%
435	├─	445	1	0,4%	1	0,2%	2	0,3%	274	98,2%	440	99,3%	714	98,9%
445	├─	455	2	0,7%	0	0,0%	2	0,3%	276	98,9%	440	99,3%	716	99,2%
455	├─	465	2	0,7%	1	0,2%	3	0,4%	278	99,6%	441	99,5%	719	99,6%
465	├─	475	0	0,0%	1	0,2%	1	0,1%	278	99,6%	442	99,8%	720	99,7%
475	├─	485	0	0,0%	1	0,2%	1	0,1%	278	99,6%	443	100,0%	721	99,9%
485	├─	495	1	0,4%	0	0,0%	1	0,1%	279	100,0%	443	100,0%	722	100,0%
Total			279	100,0%	443	100,0%	722	100,0%						

- Construção do histograma:

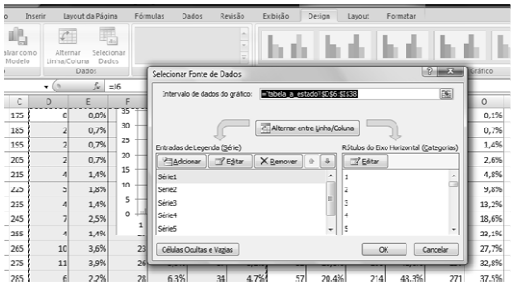
√ Selecionar as frequências na tabela montada



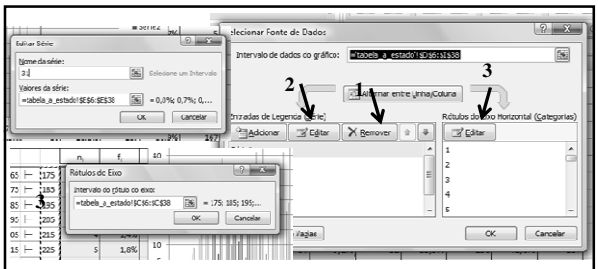
√ Clique Inserir > Colunas > Colunas 2D



√ Selecione o gráfico e clique em Selecionar Dados (guia Ferramentas de Gráfico – Design)



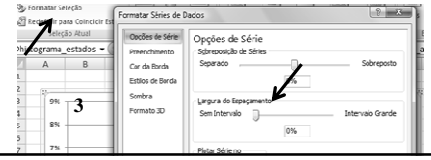
√ Utilize o tutorial para 'limpar' seu gráfico

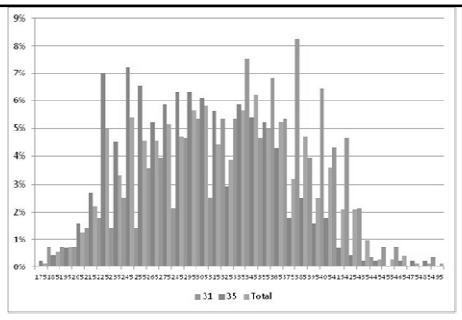


1. Remover séries (deixar apenas as de frequência relativa)
2. Editar as séries restantes e nomeá-las
3. Editar os rótulos do eixo horizontal (eixo *x*)

- Na guia Layout, aprimore seu gráfico:

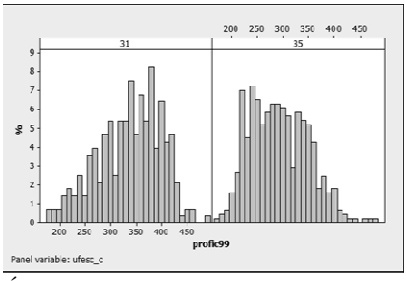
1. Edite as legendas (se desejar)
2. Edite os eixos dos gráficos
3. Selecione as barras e formate-as
   - √ Deixar as barras contíguas (s/ intervalo de espaçamento)





- Qual a forma da distribuição do desempenho?
  - √ Compare-as por estado
  - √ Compare-as em relação ao global

- Uma outra visão



√ É mais 'limpa'?

√ Facilitou a análise?

√ As conclusões são as mesmas?

Item 2 – Influência Gênero e Etnia

- Questão:
  - √ Há evidências empíricas que suportem a hipótese de que gênero e etnia influenciem o desempenho?
- Objetivo:
  - √ Visualizar graficamente possíveis diferenças (ou igualdades ➔ diferença = 0)

- Tabela de medidas-resumo - Etnias:

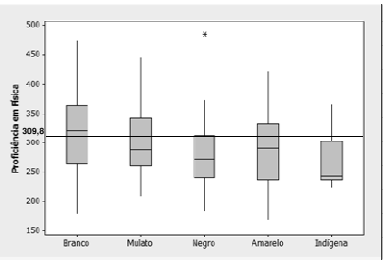
	Global	Etnia				
		Branco	Mulato	Negro	Amarelo	Indígena
Média	309,807	315,526	302,057	281,773	288,606	267,673
Desvio-padrão	62,414	63,118	57,198	62,768	57,989	46,278
Coeficiente de variação	20,146	20,004	18,936	22,276	20,093	17,289
Mínimo	168,280	177,200	208,520	183,760	168,280	224,210
Máximo	485,040	475,910	446,330	485,040	422,050	365,440
Quantidade (n)	722	509	145	23	36	9

√ Há evidências que indiquem diferenças de desempenho entre alunos de diferentes etnias? As diferenças são significativas?

√ Quais suas conclusões?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Desempenho e Etnia



√ Há evidências de diferenças no desempenho médio?

√ Há evidências de diferenças na dispersão?

- Qual seu significado?

√ Há *outliers*? Quais?

- Tabela de medidas-resumo - Sexo:

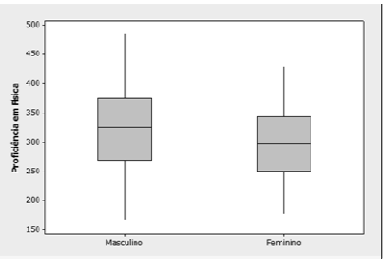
	Global	Sexo	
		Masculino	Feminino
Média	309,807	322,492	300,516
Desvio-padrão	62,414	65,267	58,336
Coeficiente de variação	20,146	20,238	19,425
Mínimo	168,280	168,280	177,200
Máximo	485,940	485,940	429,820
Quantidade (n)	722	309	413

√ Há evidências que indiquem diferenças de desempenho entre alunos do sexo masculino e feminino? Elas são significativas?

√ Quais suas conclusões?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Desempenho e Gênero

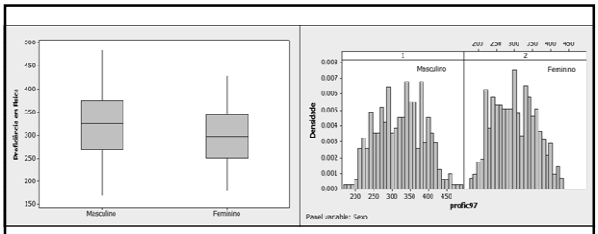


√ Há evidências de diferenças no desempenho médio?

√ Há evidências de diferenças na dispersão?

- Qual seu significado?

√ Há *outliers*? Quais?

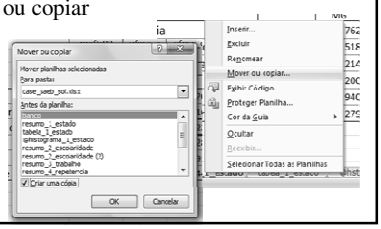


- Neste caso, o histograma facilitou sua análise?
  - √ Suas conclusões são as mesmas?

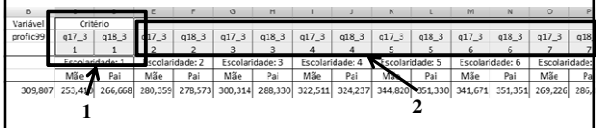
Item 3 – Influência Escolaridade Pais

- Questão:
  - √ O nível de escolaridade dos pais ou das mães influenciam mais no desempenho? Ou é indiferente?
- Objetivo:
  - √ Utilizar médias (e desvio-padrão) para observar evidências de possíveis diferenças (ou indiferente ➔ diferença = 0)

- Montar uma tabela para cálculo de:
  - √ média
  - √ desvio padrão
  - √ coeficiente de variação
- Faça uma cópia da guia *resumo_1_estado*:
  - √ Com o cursor na guia da planilha que deseja copiar, clique em Mover ou copiar
  - √ Marque copiar



- Na nova guia:
  1. Mude o critério de estratificação
  2. Copie o grupo de colunas 5 vezes e substitua valor pelos níveis de escolaridade desejados
     2; 3; 4; 5; 6 e 7



Variável	Critério														
profic99	q17_3 1	q18_3 1	q17_3 2	q18_3 2	q17_3 3	q18_3 3	q17_3 4	q18_3 4	q17_3 5	q18_3 5	q17_3 6	q18_3 6	q17_3 7	q18_3 7
	Escolaridade: 1		Escolaridade: 2		Escolaridade: 3		Escolaridade: 4		Escolaridade: 5		Escolaridade: 6		Escolaridade: 7	
	Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai
309,807	253,418	266,668	280,359	278,573	300,314	288,330	322,511	324,237	344,820	351,330	341,671	351,351	269,226	286,

1 2

- Tabela de medidas-resumo:

	Variável	Critério													
	profic99	q17_3 1	q18_3 1	q17_3 2	q18_3 2	q17_3 3	q18_3 3	q17_3 4	q18_3 4	q17_3 5	q18_3 5	q17_3 6	q18_3 6	q17_3 7	q18_3 7
		Escolaridade: 1		Escolaridade: 2		Escolaridade: 3		Escolaridade: 4		Escolaridade: 5		Escolaridade: 6		Escolaridade: 7	
		Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai	Mãe	Pai
Média		309,807	253,418	266,668	280,359	278,573	300,314	288,330	322,511	324,237	344,820	351,330	341,671	351,351	269,226	286,1
Desvio-padrão		62,414	39,021	45,841	53,819	52,487	69,043	56,928	54,996	54,859	56,944	56,740	60,097	49,732	51,723	56,0
Coeficiente de variação		20,146	15,398	18,690	19,018	18,841	19,993	19,432	17,052	16,919	16,403	16,150	17,706	14,148	19,212	19,6
Mínimo		168,280	168,280	168,280	178,210	181,360	177,200	177,200	183,760	194,230	194,230	195,400	205,380	211,150	212,030	216,5
Máximo		485,940	336,910	420,830	458,520	420,710	446,330	446,330	453,870	459,490	485,940	485,940	469,860	435,660	435,660	431,9
Quantidade (n)		722	24	30	191	193	169	123	148	175	177	155	40	32	53	

√ Considerado o nível de escolaridade, há diferenças nos desempenhos? São significativas?

√ Essa diferença mantém-se a mesma para todos os 7 níveis de escolaridade considerados?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Outra estrutura da tabela:

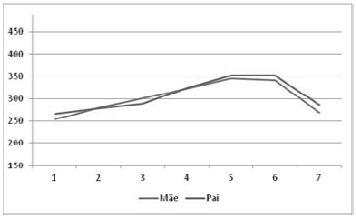
	Global	Mãe							Pai						
		1	2	3	4	5	6	7	1	2	3	4	5	6	7
Média	309,807	253,418	266,668	280,359	278,573	300,314	288,330	322,511	324,237	344,820	351,330	341,671	351,351	269,226	286,47
Desvio-padrão	62,414	39,021	45,841	53,819	52,487	69,043	56,928	54,996	54,859	56,944	56,740	60,097	49,732	51,723	56,08
Coeficiente de variação	20,146	15,398	18,690	19,018	18,841	19,993	19,432	17,052	16,919	16,403	16,150	17,706	14,148	19,212	19,62
Mínimo	168,280	168,280	168,280	178,210	181,360	177,200	177,200	183,760	194,230	194,230	195,400	205,380	211,150	212,030	216,54
Máximo	485,940	336,910	420,830	458,520	420,710	446,330	446,330	453,870	459,490	485,940	485,940	469,860	435,660	435,660	431,90
Quantidade (n)	722	24	30	191	193	169	123	148	175	177	155	40	32	53	

√ Sua análise ficou facilitada?

√ Suas conclusões são as mesmas?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Gráfico das médias de desempenho por nível de escolaridade (pai/mãe)

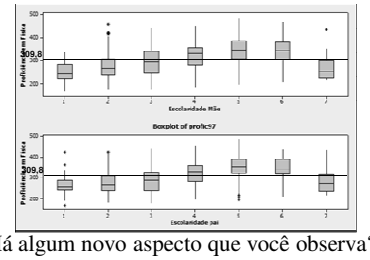


√ Facilitou sua análise?

√ Suas conclusões são as mesmas?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Box-plot do desempenho por nível de escolaridade pais



√ Há algum novo aspecto que você observa?

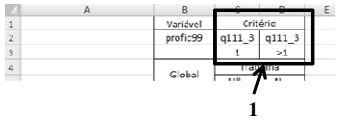
√ Suas conclusões são as mesmas?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

Item 4 – Influência do Trabalho

- Questão:
  - √ Há evidências na influência do trabalho do aluno em seu desempenho?
- Objetivo:
  - √ Visualizar possível diferença por meio de medidas resumo (ou graficamente)
    - igualdade desempenho → diferença = 0 → não influencia

- Montar uma tabela para cálculo de:
  - √ média
  - √ desvio padrão
  - √ coeficiente de variação
- Faça uma cópia da guia *resumo_1_estado*:
  1. Mude o critério de estratificação



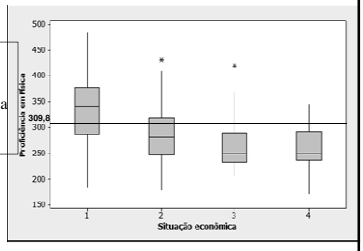
- Tabela de medidas-resumo:

	Global	Trabalha	
		Não	Sim
Média	309,807	331,473	277,717
Desvio-padrão	62,414	60,801	49,513
Coeficiente de variação	20,146	18,343	17,829
Mínimo	168,280	183,760	168,280
Máximo	485,940	485,940	431,460
Quantidade (n)	722	431	291

- √ Considerado situação econômica, há diferença significativa nos desempenhos?
- √ Quais suas conclusões?
- √ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Box-plot do desempenho por situação trabalhista

1. Não trabalha
2. Trabalha e depende família
3. Trabalha e não depende família
4. Trabalha e família depende

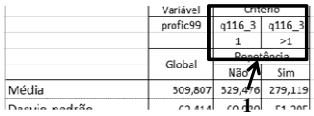


- √ Há algum novo aspecto que você observa?
- √ Suas conclusões são as mesmas?
- √ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

Item 5 – Influência Repetência

- Questão:
  - √ Há evidências da influência da repetência no desempenho do aluno em Física?
- Objetivo:
  - √ Visualizar possível diferença por meio de medidas resumo (ou graficamente)
    - igualdade desempenho ➔ diferença = 0 ➔ não influencia

- Montar uma tabela para cálculo de:
  - √ média
  - √ desvio padrão
  - √ coeficiente de variação
- Faça uma cópia da guia *resumo_1_estado*:
  1. Mude o critério de estratificação



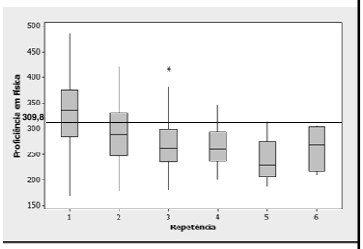
- Tabela de medidas-resumo:

	Global	Repetência	
		Não	Sim
Média	309,807	329,476	279,119
Desvio-padrão	62,414	60,930	51,205
Coeficiente de variação	20.146	18,493	18,345
Mínimo	168,280	168,280	177,200
Máximo	485,940	485,940	420,830
Quantidade (n)	722	440	282

  - √ Considerada repetências no passado, há evidências de influência no desempenho? Ela é significativa?
  - √ Quais suas conclusões?
  - √ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Box-plot do desempenho por situação trabalhista

1. Não
2. Sim, 1 vez
3. Sim, 2 vezes
4. Sim, 3 vezes
5. Sim, 4 vezes
6. Sim, mais de 4 vezes



√ Há algum novo aspecto que você observa?

√ Suas conclusões são as mesmas?

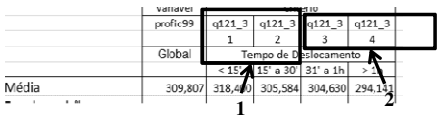
√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

Item 6 – Local de Residência

- Questão:
  - √ Há evidências de que os alunos que moram mais próximos da escola tenham melhor desempenho em Física?
    - Para essa particular amostra de alunos
- Objetivo:
  - √ Visualizar possível diferença por meio de medidas resumo (ou graficamente)
    - igualdade desempenho → diferença = 0 → não influencia

- Montar uma tabela para cálculo de:
  - √ média
  - √ desvio padrão
  - √ coeficiente de variação
- Faça uma cópia da guia *resumo_1_estado*:
  1. Mude o critério de estratificação
  2. Copie o grupo de colunas 1 vez e substitua valor pelos níveis de tempo de deslocamento desejados

3 e 4



- Tabela de medidas-resumo:

	Global	Tempo de Deslocamento			
		< 15'	15' a 30'	31' a 1h	> 1h
Média	309,807	318,400	305,584	304,630	254,141
Desvio-padrão	62,414	65,428	61,538	54,498	65,719
Coeficiente de variação	20,146	20,549	20,138	17,890	22,343
Mínimo	168,280	186,930	177,200	202,880	168,280
Máximo	485,940	485,940	458,520	431,460	413,350
Quantidade (n)	722	270	296	131	25

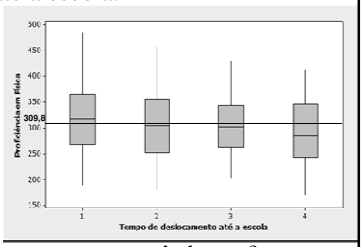
√ Considerado tempo de deslocamento, você encontra evidências que suportem sua influência no desempenho do aluno? A influência é significativa?

√ Quais suas conclusões?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Box-plot do desempenho por tempo de deslocamento até a escola:

1. Menos de 15 minutos
2. De 15 a 30 minutos
3. De 31 minutos a 1 hora
4. Mais de 1 hora



√ Há algum novo aspecto que você observa?

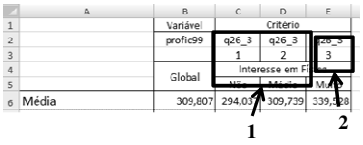
√ Suas conclusões são as mesmas?

√ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

Item 7 – Influência Interesse Pessoal

- Questão:
  - √ Alunos que gostam de Física parecem ter um desempenho melhor do que aqueles que não gostam?
- Objetivo:
  - √ Visualizar possível diferença por meio de medidas resumo (ou graficamente)
    - igualdade desempenho ➔ diferença = 0 ➔ não influencia

- Montar uma tabela para cálculo de:
  - √ média
  - √ desvio padrão
  - √ coeficiente de variação
- Faça uma cópia da guia *resumo_1_estado*:
  1. Mude o critério de estratificação
  2. Copie uma coluna substitua o valor pelos nível de interesse em Física desejado (3)



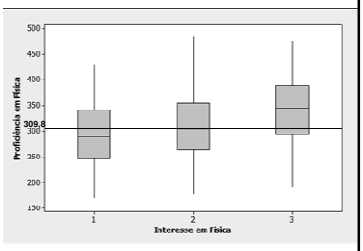
- Tabela de medidas-resumo:

	Global	Interesse em Física		
		Não	Médio	Muito
Média	309,807	294,037	309,739	339,528
Desvio-padrão	62,414	57,238	60,160	66,912
Coeficiente de variação	20,146	19,466	19,423	19,707
Mínimo	168,280	168,280	177,200	189,670
Máximo	485,940	429,820	485,940	475,910
Quantidade (n)	722	234	363	125

  - √ Há evidências que indiquem diferenças nos desempenhos entre os três grupos de alunos? Elas são significativas?
  - √ Quais suas conclusões?
  - √ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

- Box-plot do desempenho por interesse em Física:

1. Não gosto
2. Gosto mais ou menos
3. Gosto muito



  - √ Há algum novo aspecto que você observa?
  - √ Suas conclusões são as mesmas?
  - √ Que hipótese(s) você gostaria de testar de uma maneira mais formal?

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