

# Estatística descritiva / Análise Exploratória

última atualização: 5 de junho de 2012

1. Foram feitas medições dos teores de um poluente em duas regiões ( $A$  e  $B$ ), representadas nos gráficos da figura a seguir.

- Indique qual *boxplot* da figura à direita correspondente cada curva da figura à esquerda. Justifique sua resposta.
- Em uma das regiões a média foi de 44.6 e a mediana 40.6, enquanto que em outra a média foi 49.5 e a mediana 49.2. Quais valores correspondem a cada região? Justifique sua resposta.
- Interprete e discuta cada um dos gráficos, comparando as regiões.

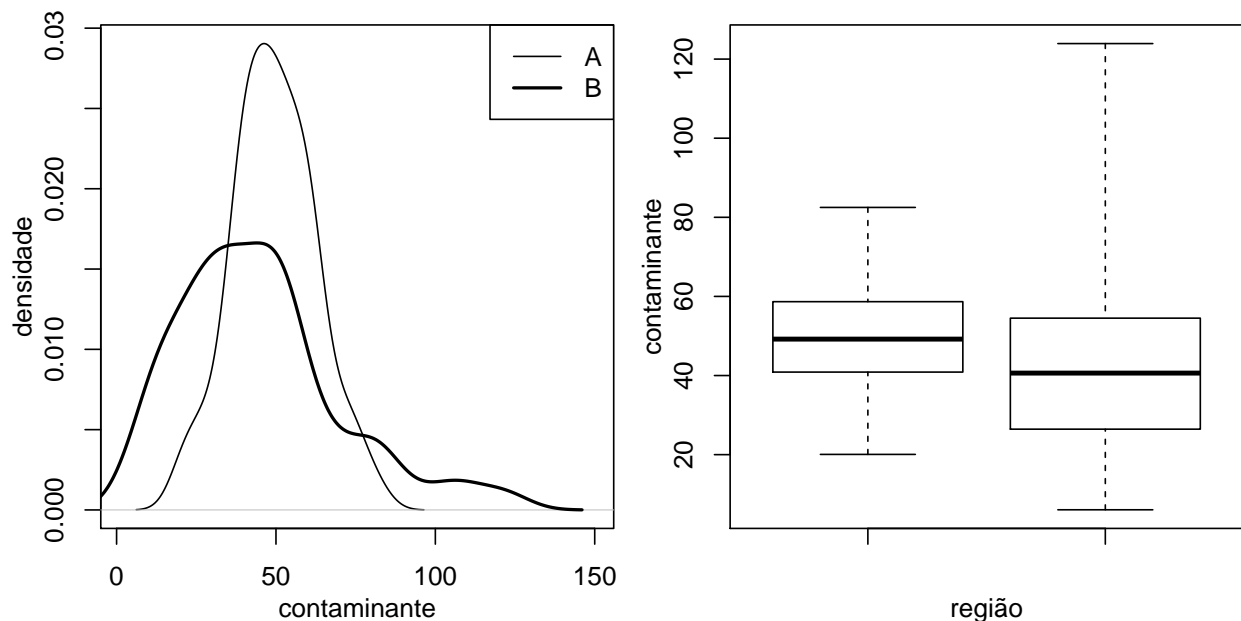


Figura 1: Teores de poluente medidos em amostras tomadas em duas regiões.

2. Foram feitas medições de índices de qualidade da água em 20 locais e os dados coletados foram:

89.6 86.2 49.0 82.4 81.5 76.2 94.8 90.7 88.5 77.3

81.8 89.5 75.6 97.8 71.6 88.7 93.6 86.0 93.3 91.1

- faça um histograma dos dados
- faça um diagrama ramo-e-folhas
- faça um gráfico *boxplot*
- obtenha a média e desvio padrão
- obtenha o coeficiente de variação
- obtenha a amplitude e a amplitude interquartilica

(g) caracterize a distribuição dos dados

3. Um estudo procurou relacionar medidas de um índice de poluição (PM10) com atendimentos hospitalares por doenças respiratórias. Foram anotados dados em vários períodos e em cinco capitais.

Discuta estratégias para investigar a relação desejada a partir dos dados. Mencione que tipos de análises estatísticas descritivas poderiam ser feitas, os possíveis cenários (resultados) e como seriam interpretados. Comente sobre o que deveria ser levado em consideração nas análises.

4. Foi feita uma pesquisa sobre as condições salariais de professores de um certo estado. Os dados foram organizados em uma tabela. A seguir é mostrada uma pequena fração dos dados. A descrição dos atributos anotados está na Tabela abaixo.

	Degree	Rank	Sex	Year	YSdeg	Salary
1	1	3	0	25	35	36350
2	1	3	0	13	22	35350
3	1	3	0	10	23	28200
4	1	3	1	7	27	26775
5	0	3	0	19	30	33696
6	1	3	0	16	21	28516

Atributo	Descrição
Degree	Formação: 1: Doutorado, 0: Mestrado
Rank	Cargo (1: Prof Assistente, 2: Prof Associado, 3: Prof Pleno)
Sex	1: feminino, 0: masculino
Year	Anos de trabalho
YSdeg	Anos desde a obtenção da maior titulação
Salary	Salário em dolares por ano

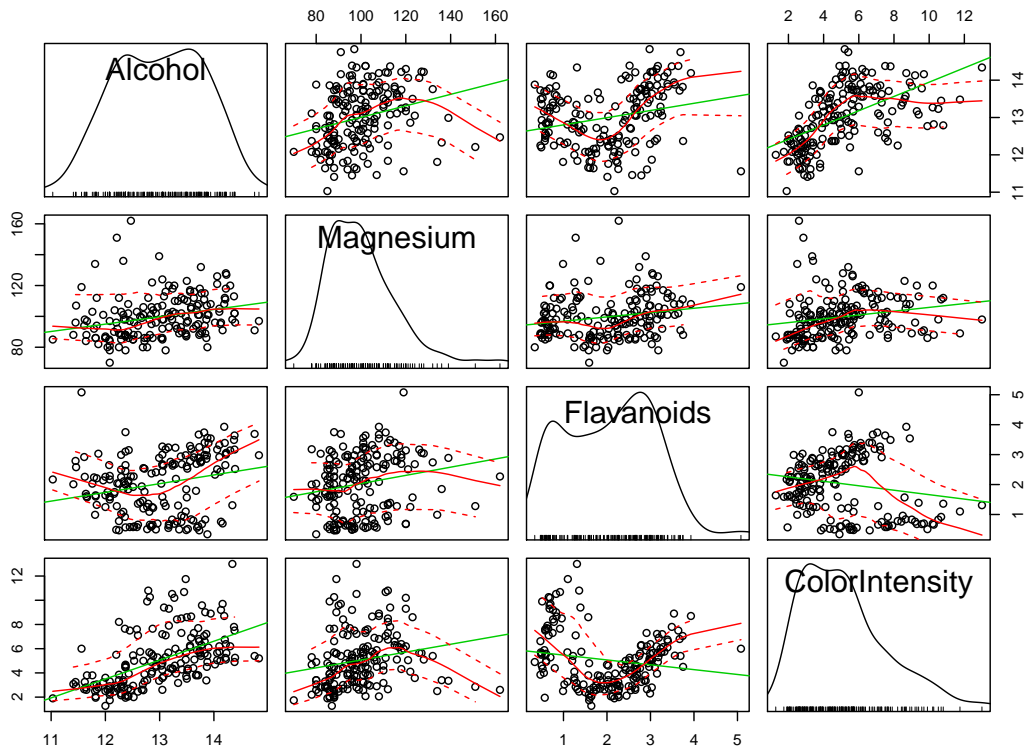
- (a) Classifique cada um dos atributos (variáveis).
- (b) Esboce um gráfico adequado para resumir cada um dos atributos individualmente
- (c) Como voce investigaria (por exemplo, que tipo de gráfico) se existe relação entre:
- Sexo e Formação
  - Sexo e Salário
  - Anos de trabalho e salário
5. A tabela a seguir apresenta as notas de matemática no vestibular e na disciplina de cálculo de alguns alunos selecionados ao acaso. Pretende-se examinar os desempenhos nestas provas e se há relação entre os desempenhos.

Aluno	Vestibular	Cálculo	Aluno	Vestibular	Cálculo
1	37	65	7	35	50
2	57	92	8	80	90
3	34	56	9	65	88
4	40	70	10	47	71
5	21	52	11	28	52
6	28	73	12	67	88

- (a) Calcule a mediana, quartis e amplitude interquartílica das notas de cálculo.
- (b) Calcule o coeficiente de variação das notas do vestibular e de cálculo.
- (c) Construa um diagrama "ramo-e-folhas" com todas as notas (vestibular e cálculo) e marque (sublinhe) nas "folhas" os dados da prova de cálculo.
- (d) Faça um gráfico com os diagramas "box-plot" das duas notas (um "boxplot" para cada).
- (e) Construa um gráfico adequado para representar os dados das duas provas conjuntamente. Calcule medida(s) de associação adequada(s).

(f) Compare, interprete e discuta os resultados.

6. Uma série de características químicas foram medidas em diferentes vinhos. Os gráficos a seguir mostram quatro delas. Discuta os gráficos e suas interpretações utilizando conceitos e princípios de análise estatística descritiva/exploratória de dados. Inclua na sua discussão possíveis tratamentos dos dados.



7. Uma empresa que tem 200 funcionários paga um salário médio de R\$ 800,00. Está sendo estudada uma política de reajuste salarial e duas propostas foram apresentadas: a) aumento de 15% para todos os funcionários ou b) aumento de R\$ 120,00 para todos os funcionários. Avalie e discuta as propostas e suas consequências para a empresa e para os funcionários. Utilize medidas estatísticas para fundamentar os seus argumentos.

8. Alunos de um centro acadêmico resolveram fazer uma pesquisa sobre uso e percepção de transporte e acessibilidade dos alunos ao campus. Os dados disponíveis foram organizados em uma planilha com linhas e colunas. Cada linha contém os dados de um aluno e cada coluna os atributos pesquisados. Foi tomada uma amostra de 275 alunos e uma pequena parte dos dados (três alunos) é mostrada a seguir para ilustrar a estrutura do conjunto de dados.

Setor	Sexo	Ano	Tr	Prop	Proc	Opi	TempoI	TempoV
Ex	M	2010	On	0	CTBA	MB	35	20
Tec	F	2009	B	1	Fora	B	10	10
Bio	F	2010	C	0	Fora	Reg	20	15
...								

Os atributos são:

- *Setor* : setor do curso dos aluno (exatas, tecnológica, ciências sociais, etc)
- *Sexo* : sexo do(a) aluno(a) (M: masculino, F: feminino)
- *Ano* : ano de ingresso na universidade
- *Tr* : meio de transporte mais utilizado (Obibus, Bicicleta, carro, etc)
- *Proc* : procedência (Curitiba ou fora)

- *Opi* : opinião sobre o transporte (MuitoBom, Bom, Regular, Ruim, MuitoRuim)
- *TempoI* : tempo tipicamente gasto para ir ao Campus
- *TempoV* : tempo tipicamente gasto para voltar do Campus

Considerando estes dados disponíveis, responda *baseando-se em possíveis análises estatísticas*.

- Classifique cada uma das variáveis.
- Se voce tivesse que resumir cada um dos atributos (variável) por alguma medida estatística, qual medida voce calcularia para cada um deles?
- Considere agora que voce faria um gráfico para ilustrar cada uma das variáveis. Esboce o gráfico que seria utilizado para cada variável.
- Deseja-se investigar se o *Tempo* está associado ao tipo de transporte. Como voce verificaria se existe esta relação?
- Como verificar através de análises dos dados se o tempo de ida e volta estão associados?
- Qual(ais) medida(s) estatística(s) poderia(m) ser calculada(s) para verificar se as opiniões sobre o transporte diferem em relação aos sexos dos alunos?
- Formule pelo menos mais quatro questões a serem investigadas com os dados dizendo como seriam investigadas com as análises.

9. Considere os dados a seguir:

35 39 50 47 33 17 17 80 23 51 12 30 20  
18 20 21 20 45 12 81 98 47 34 25 19 27

- Calcule a média e mediana.
  - Obtenha os quartis.
  - Obtenha o 1º e o 9º decil (isto é, os quantis 0,10 e 0,90).
  - Calcule o desvio padrão e coeficiente de variação.
  - Qual seria o valor de cada uma das medidas acima se:
    - acrescente-se 15 a cada dado,
    - cada dado fosse aumentado em 10%.
  - Obtenha uma distribuição de frequência para os dados.
  - Obtenha o histograma dos dados.
  - Desenhe o diagrama ramo-e-folhas.
  - Faça o box-plot dos dados
10. Considere agora que os dados da questão anterior são referentes a dois grupos. Os da primeira linha são do grupo *A* e os da segunda do grupo *B*. Compare os grupos utilizando análise estatísticas dos dados.
11. Um professor selecionou aleatoriamente 15 alunos. Os alunos foram classificados considerando frequência em 'menor' que 80% e frequência 'maior' que 80%. Além disso, considerou o sexo e a nota na primeira avaliação. Os dados desses 15 alunos estão abaixo:

Sexo	Presença	Nota
Fem	>80%	79
Masc	>80%	69
Masc	>80%	69
Masc	<80%	28
Masc	>80%	83
Fem	<80%	76
Masc	>80%	69
Fem	>80%	38
Masc	<80%	100
Masc	>80%	97
Masc	<80%	48
Masc	<80%	62
Fem	>80%	76
Fem	>80%	72
Masc	<80%	28

- (a) Obtenha uma distribuição de frequências de cada uma das variáveis
- (b) Faça uma visualização gráfica adequada de cada uma das distribuições de frequências

12. Considere a variável **Nota** dos dados do exercício anterior

- (a) Faça um box-plot
- (b) Faça um ramo-e-folhas
- (c) Calcule a média, variância e coeficiente de variação.

- (a)
- (b) Faça um ramo-e-folhas
- (c) Calcule a média, variância e coeficiente de variação.

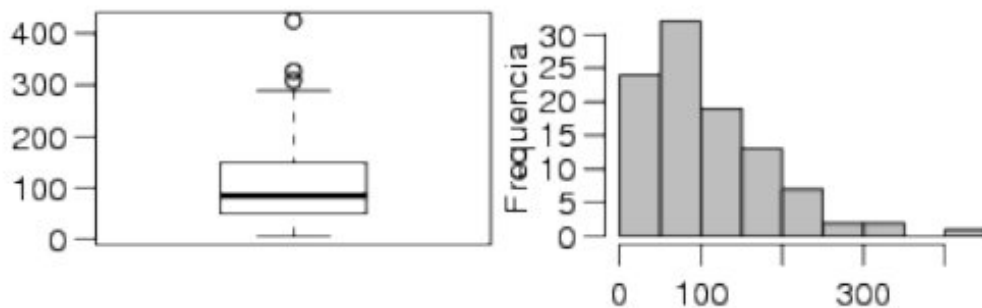
13. Considere as variáveis **Sexo** e **Presença**

- (a) Você acha que há associação entre essas duas variáveis? Justifique calculando uma medida de associação e fazendo um gráfico adequado.

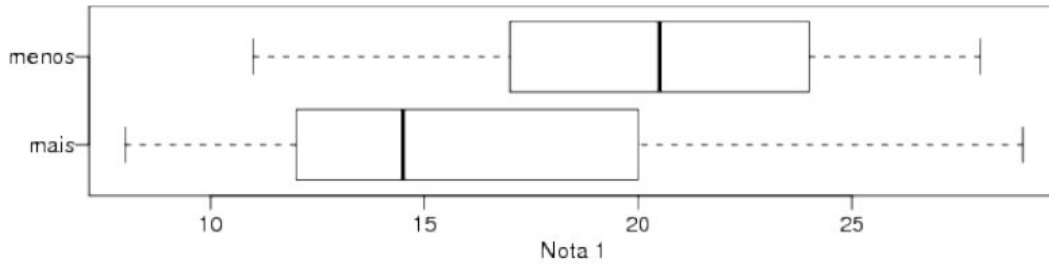
14. Considere as variáveis **Presença** e **Nota**

- (a) Faça também um gráfico adequado com as duas variáveis.
- (b) Você acha que há associação entre essas duas variáveis? Justifique calculando medida estatísticas adequadas.

15. Considere o histograma e o box-plot a seguir. Comente sobre esses gráficos. A média é maior ou menor que a mediana nesse caso?



16. Considere que um professor selecionou um grupo de alunos e avaliou a nota da primeira prova (cujo valor era 30pts) e a frequência às aulas. Os alunos foram classificados considerando frequência em 'menos' que 80% e frequência em 'mais' que 80%. Após isso, foi construído um box-plot da nota para cada um desses dois grupos. Que se pode concluir analisando os box-plots?



17. Considerando o item anterior, os alunos foram também classificados quanto à nota em dois grupos: maior que 70% e menor que 70% da nota máxima. Após isso, fez-se uma tabela da classificação por frequência do item anterior versus essa classificação por nota. A tabela de 30 alunos sorteados aleatoriamente está apresentada a seguir:

		nota	
presenca	<70%	>70%	
<80%	7	4	
>80%	5	14	

- (a) Calcule uma medida de associação  
 (b) Você acha que há associação entre essas duas variáveis? Justifique
18. Considerando os 30 alunos do item anterior, a variância da nota desses 30 alunos é 32 e a variância desses 30 alunos com menos e mais de 80% de frequência é 39 e 23 respectivamente. Calcule o percentual da variabilidade de notas que é explicada pela classificação quanto à presença.

19. Uma pesquisa considerou 102 ocupações (profissões) para as quais foram anotados dados de: EDU (número de anos de educação); Rend) renda; FEM (% de mulheres na profissão); PREST (*escore* que mede o prestígio da profissão e TIPO (bc: trabalho com uniforme, em geral manual; wc: trabalho em uniforme formal e prof: profissional/gerencial/técnico). Foram feitos alguns análises e os resultados são mostrados a seguir. Interprete e discuta os resultados.

