

# UNIVERSIDADE FEDERAL DO ABC

## Lista 8 - Introdução à Probabilidade e Estatística

### Estatística Descritiva

- 1** — Defina amostra e população.
- 2** — Defina parâmetro e estatística.
- 3** — Classifique cada uma das variáveis seguintes como quantitativa, qualitativa e como contínua e discreta.
- a) A concentração de impurezas em uma amostra de leite, em mg por litro.
  - b) O partido político que um indivíduo votou.
  - c) A procedência de cada ingressante da UFABC em certo ano.
  - d) O tempo de reação de um indivíduo após submetido a certo estímulo.
  - e) O número de moradores em cada residência de uma cidade.
  - f) A temperatura em certa região, em determinada época do ano.
  - g) A produção por hectare de determinado tipo de grão.
  - h) A gravidade de um fermento.
- 4** — Durante certo período de tempo as taxas de juros para dez ações foram as abaixo registradas:

Ação	01	02	03	04	05	06	07	08	09	10
Taxa %	2,59	2,64	2,60	2,62	2,55	2,61	2,50	2,63	2,64	2,69

Calcule:

- a) a taxa média;
  - b) a taxa mediana;
  - c) a taxa modal;
  - d) o desvio padrão das taxas;
  - e) o coeficiente de variação das taxas.
- 5** — Os dados a seguir foram obtidos em indivíduos contaminados pelo veneno de um certo tipo de inseto e submetidos a tratamento. A variável de interesse  $R$  é definida como o tempo (em horas) entre a administração do tratamento e a recuperação do indivíduo. Os valores de  $R$  são os seguintes: 3, 90, 23, 46, 2, 42, 47, 37, 12, 51, 11, 1, 3, 3, 45, 3, 4, 11, 2, 8, 56, 39, 22, 16, 5 e 52.
- a) Determine a média, mediana, intervalo inter-quartil e desvio padrão.
  - b) Separe o conjunto de dados em três grupos denominados *cura rápida*, com valor de  $R$  menor ou igual a 12, *cura normal*, se o valor de  $R$  for maior do que 12 e menor ou igual a 45, e *cura lenta*, se o valor de  $R$  estiver acima de 45. Compare a variabilidade desses três grupos através do coeficiente de variação desses grupos.

**6** — Um órgão do governo do estado está interessado em determinar padrões sobre o investimento em educação, por habitante, realizado pelas prefeituras. De um levantamento de dez cidades, foram obtidos os valores (codificados) da tabela abaixo:

Cidade	A	B	C	D	E	F	G	H	I	J
Investimento	25	16	14	10	19	15	19	16	19	18

- Calcule a média das observações.
- Receberão um programa especial as cidades com valores de investimento inferiores à média menos o desvio padrão. Alguma cidade receberá o programa

**7** — As concentrações de óxido de nitrogênio e hidrocarbono (em  $\mu\text{g}/\text{m}^3$ ) foram determinadas em uma área urbana, em locais e horários específicos. Os dados são mostrados a seguir.

Dia	Óxido de Nitrogênio (O)	Hidrocarbono (H)	DIF = O - H
1	104	108	-4
2	116	118	-2
3	84	89	-5
4	77	71	6
5	61	66	-5
6	84	83	1
7	81	88	-7
8	72	76	-4
9	61	68	-7
10	97	96	1
11	84	81	3

- Classifique as variáveis em estudo.
- Realize uma análise descritiva dos dados. Calcule média e desvio padrão para cada variável e para a variável  $\text{DIF} = \text{O} - \text{H}$  (diferença entre as concentrações dos poluentes).
- Considerando a variável DIF, pode-se dizer que as duas classes de poluentes estão presentes nas mesmas concentrações?

**8** — Em uma granja foi observado a distribuição dos frangos em relação ao peso, que era o seguinte:

Peso (gramas)	número de frangos
960[-980	60
980[-1.000	160
1.000[-1.020	280
1.020[-1.040	260
1.040[-1.060	160
1.060 [- 1.080	80

- Qual a media da distribuição?
- Qual a variância da distribuição?
- Construa o histograma e comente os resultados encontrados.
- Queremos dividir os frangos em quatro categorias, em relação ao peso, de modo que:
  - os 20% mas leves sejam da categoria D
  - os 30% seguintes sejam da categoria C

- c) os 30% seguintes sejam da categoria B
- d) os 20% mais pesados sejam da categoria A

Determine quais são os limites de peso entre as categorias A, B, C e D. Faça uma representação gráfica dos resultados obtidos.

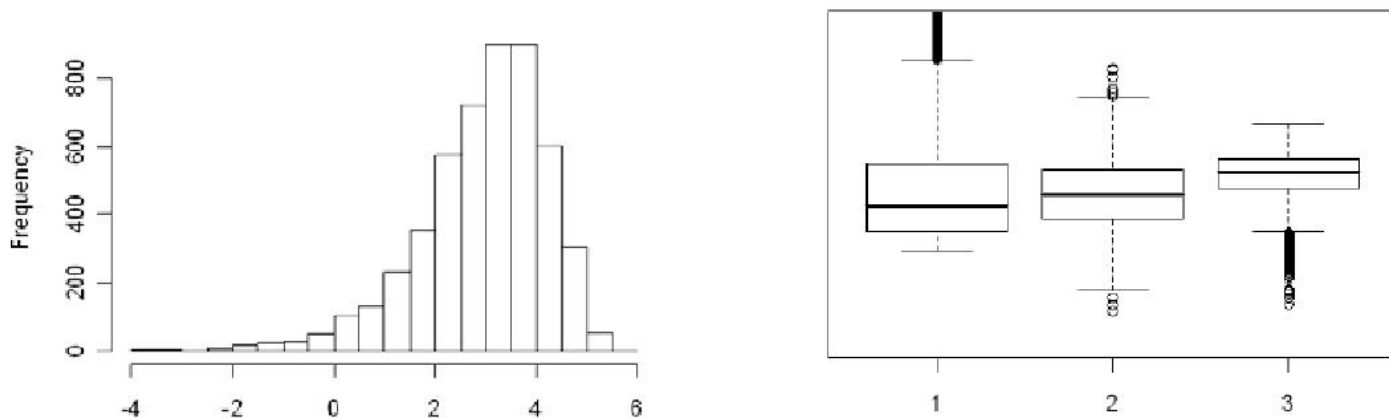
- e) O granjeiro decide separar deste lote os animais com peso inferior a dois desvios padrões abaixo da média para receberem ração reforçada, e também separar os animais com peso superior a um e meio desvio padrão acima da média para usa-los como reprodutores. Qual a porcentagem de animais que serão separados em cada caso?

**9 —** Uma companhia distribuidora tem por hipótese que uma chamada telefônica é mais eficiente que uma carta para acelerar a cobrança de contas atrasadas. Esta companhia fez uma experiência usando duas amostras e obteve os seguintes resultados:

Método utilizado	Nº de dias até o pagamento						
Carta	10	8	9	11	11	14	10
Chamada telefônica	7	4	5	4	8	6	9

Qual dos métodos apresentou resultados mais homogêneos? Justifique através do coeficiente de variação.

**10 —** É dado na Fig. 1 o histograma para um conjunto de dados. Qual dos *box-plots* é consistente com o histograma na descrição da distribuição dos dados. Justifique sua escolha.



Histograma e box-plots

**11 —** Considere o seguinte conjunto de dados de uma variável numérica:

21	21	21	22	22	23	23	23	24	24
25	25	25	25	26	26	26	28	30	31
31	32	33	33	33	34	34	35	35	36

- a) Calcule as medidas de posição (média, moda, mediana, primeiro e terceiro quartil) e dispersão (amplitude, variância, desvio-padrão) para os dados.
- b) Faça o histograma, e Box-plot para os dados.

- c) Comente sobre os formatos dos dados (assimetria, posição, variabilidade, modalidade (modas)), além da presença ou não de outliers.
- d) Para os dados agrupados na tabela de frequências do histograma, calcule média, moda, mediana, variância e desvio-padrão. Média, mediana, variância e desvio-padrão são próximos dos dados não-agrupados (obtidos no item a)).

**12** — Os dados a seguir correspondem aos recordes de atletas de 10 países na Olimpíada de los Angeles em 1984 em algumas provas de atletismo.

#### Mulheres

País	100m (seg)	400m (seg)	3000m (min)	Maratona (min)
Argentina	11,61	54,50	9,79	178,52
Brasil	11,31	52,80	9,77	168,75
Chile	12,00	54,90	9,37	171,38
Colômbia	11,6	53,26	9,46	165,42
Alemanha	11,01	48,16	8,75	148,53
França	11,15	51,73	8,98	155,27
Portugal	11,81	54,30	8,84	151,20
Canadá	11,00	50,06	8,81	149,50
USA	10,79	50,62	8,50	142,72
Kenya	11,73	52,70	9,20	181,05

#### Homens

País	100m (seg)	400m (seg)	3000m (min)	Maratona (min)
Argentina	10,39	46,84	14,04	137,72
Brasil	10,22	45,21	13,62	133,13
Chile	10,34	46,20	13,61	134,03
Colômbia	10,43	46,10	13,49	131,35
Alemanha	10,16	44,50	13,21	132,23
França	10,11	45,28	13,34	132,30
Portugal	10,53	46,70	13,13	128,65
Canadá	10,17	45,68	13,55	131,15
USA	9,93	43,86	13,20	128,22
Kenya	10,46	44,92	13,10	129,75

- a) Para cada gênero e modalidade calcule a média, mediana e desvio padrão dos recordes.
- b) Compare os resultados em a). Em qual modalidade as diferenças quanto ao gênero é maior? E em qual é menor?
- c) Utilizando gráficos boxplot faça uma comparação do desempenho entre os sexos nas quatro modalidades. Comente.

**13** — O departamento de vendas de certa companhia foi formado há um ano com a admissão de 15 vendedores. Nessa época, foram observados para cada um dos vendedores os valores de quatro variáveis:

- V: Classificação do vendedor;
- T: resultado em um teste apropriado para vendedores;
- E: anos de experiência de vendas;
- G: conceito do gerente de venda, quanto ao currículo do candidato.

O diretor da companhia resolveu agora ampliar o quadro de vendedores e pede sua colaboração para responder algumas perguntas. Para isso, ele lhe dá informações adicionais sobre duas variáveis:

- V: volume médio mensal de vendas;
- Z: zona da capital para a qual o vendedor foi designado.

O quadro de resultados é o seguinte:

Vendedores	T	E	G	V	Z
1	8	5	Bom	54	Norte
2	9	2	Bom	50	Sul
3	7	2	Mau	48	Sul
4	8	1	Mau	32	Oeste
5	6	4	Bom	30	Sul
6	8	4	Bom	30	Oeste
7	5	3	Bom	29	Norte
8	5	3	Bom	27	Norte
9	6	1	Mau	24	Oeste
10	7	3	Mau	24	Oeste
11	4	4	Bom	24	Sul
12	7	2	Mau	23	Norte
13	3	3	Mau	21	Sul
14	5	1	Mau	21	Oeste
15	3	2	Bom	16	Norte

Dados:

- $\sum T = 91$
- $\sum T^2 = 601$
- $\sum TV = 2959$
- $\sum E = 40$
- $\sum E^2 = 128$
- $\sum EV = 1260$
- $\sum V = 453$
- $\sum V^2 = 15509$

a)

- a) Classifique cada uma das variáveis da tabela quanto à escala (nominal, ordinal, intervalar e de proporção) e quanto a variável (quantitativa ou qualitativa).
- b) Faça um histograma da variável T.
- c) Qual a moda da variável T?
- d) Qual a amplitude da variável T?
- e) Faça um diagrama de dispersão entre as variáveis T e V.
- f) Calcule a média da variável V.
- g) Calcule a variância da variável V
- h) Suponha que um vendedor seja considerado excepcional se seu volume de vendas é dois desvios padrões superior à média geral. Quantos vendedores excepcionais existem na amostra? Justifique

b)

- a) Faça um box-plot da variável V associada aos vendedores que trabalham na zona sul.
- b) Faça um box-plot da variável V associada aos vendedores que trabalham na zona oeste.

- c) Os vendedores argumentam com o diretor que o critério de ordenamento dos vendedores pela variável  $V$  é injusto, pois há zonas de venda privilegiada. Com base nos itens anteriores, o que você acha dessa argumentação? Justifique.
- c)
- a) Calcule a correlação entre as variáveis  $T$  e  $V$ .
  - b) Calcule a correlação entre as variáveis  $E$  e  $V$ .
  - c) Com base nos itens anteriores, qual das variáveis  $T$  ou  $E$  é mais importante para contratar um futuro candidato a vaga de vendedor? Justifique

1.

**População** é um conjunto com todos os elementos sobre os quais serão estudados os parâmetros de estatística.

**Amostra** é um submulticonjunto com parte dos elementos da população sobre os quais serão estudados e inferidos os parâmetros de estatística.

2.

**Parâmetro** é uma característica inerente à uma população ou amostra que a descreve seguindo certas propriedades definidas.

**Estatística** é um ramo da matemática dedicado ao estudo de técnicas e métodos para resumir, analisar e interpretar dados de observações através de parâmetros e realizar inferências a partir desses dados.

3.

- a) Quantitativa e contínua.
- b) Qualitativo.
- c) Qualitativo.
- d) Quantitativo e contínuo.
- e) Quantitativo e discreto.
- f) Quantitativo e contínuo.
- g) Quantitativo e discreto.
- h) Qualitativo.

4.

$$a) \mu = \bar{T} = \frac{\sum_{i=1}^{10} T_i}{10} = \frac{2,50+2,55+2,59+2,60+2,61+2,62+2,63+2,64+2,64+2,69}{10} = \boxed{2,607\%}$$

$$b) \text{ par: } m = \frac{\left[ \frac{T_{10}}{2} + \frac{T_{\frac{10}{2}+1} \right]}{2} = \frac{2,61\% + 2,62\%}{2} = \boxed{2,615\%}$$

$$c) M = \boxed{2,64\%}$$

$$d) \sigma = \sqrt{\frac{\sum_{i=1}^{10} (T_i - \bar{T})^2}{n-1}} \approx \boxed{0,053\%}$$

$$e) c_v = \frac{\sigma}{\mu} \approx \frac{0,053\%}{2,607\%} \approx \boxed{2\%}$$

5.

$$a) \mu = \frac{1+2+2+3+3+3+3+4+5+8+11+11+12+16+22+23+37+39+42+45+46+47+51+52+56+90}{26}$$

$$\approx \boxed{24,38}$$

$$\text{par: } m = \frac{R_{26/2} + R_{26/2+1}}{2} = \frac{R_{13} + R_{13+1}}{2} = \frac{12 + 16}{2} = \boxed{14}$$

$$\begin{aligned} \text{par: } IQR = Q_3 - Q_1 &= \frac{R_{26 \cdot 3/4} + R_{26 \cdot 3/4+1}}{2} - \frac{R_{26/4} + R_{26/4+1}}{2} \\ &= \frac{R_{19,5} + R_{20,5}}{2} - \frac{R_{6,5} + R_{7,5}}{2} \\ &= \frac{(R_{19} + R_{20})/2 + (R_{20} + R_{21})/2}{2} - \frac{(R_6 + R_7)/2 + (R_7 + R_8)/2}{2} \\ &= \frac{R_{19} + 2R_{20} + R_{21}}{4} - \frac{R_6 + 2R_7 + R_8}{4} \\ &= \frac{42 + 2 \cdot 45 + 46}{4} - \frac{3 + 2 \cdot 3 + 4}{4} \\ &= 44,5 - 3,25 \\ &= \boxed{41,25} \end{aligned}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{26} (R_i - \mu)^2}{26 - 1}} \approx \boxed{23,57}$$

b)

Tipo	Cura		
	Rápida	Normal	Lenta
	$R \leq 12$	$12 < R \leq 45$	$45 < R$
Qnt.	#13	#7	#6
Ri	1, 2, 2, 3, 3, 3, 3, 4, 5, 8, 11, 11, 12	16, 22, 23, 37, 39, 42, 45	46, 47, 51, 52, 56, 90
Cv	388%	31%	45%

Percebe-se que o coeficiente de variação da cura rápida é muito maior a dos demais. Concluímos então que a cura rápida possui um tempo de recuperação muito mais disperso à cura lenta que por sua vez é um pouco mais disperso à cura normal.

6.

$$a) \mu = \frac{10+14+15+16+16+18+19+19+19+25}{10} = \boxed{17,1}$$



$$b) \sigma = \sqrt{\frac{\sum_{i=1}^{10} (I_i - \mu)^2}{10-1}} = 3,96$$

Tipo	$I \leq 17,1-3,96 = 13,14$	$I > 17,1-3,14 = 21,06$
Qnt.	#1	#9
Ri	D (10)	demais

Apenas a cidade D com valor de investimento 10 receberá o programa especial.

7.

a) b)

*Classificação por Óxido de Nitrogênio (O)*

Dia	Óxido de Nitrogênio (O)	Hidrocarbono (H)	DIF = O-H
5; 9	61	66; 68	-5; -7
8	72	76	-4
4	77	71	6
7	81	88	-7
11; 6; 3	84	81; 83; 89	-5; 1; 3
10	97	96	1
1	104	108	-4
2	116	118	-2
$\mu$	86,50		
$\sigma$	16,86		

*Classificação por Hidrocarbono (H)*

Dia	Óxido de Nitrogênio (O)	Hidrocarbono (H)	DIF = O-H
5	61	66	-5
9	61	68	-7
4	77	71	6
8	72	76	-4
11	84	81	3
6	84	83	1
7	81	88	-7
3	84	89	-5
10	97	96	1
1	104	108	-4
2	116	118	-2
	$\mu$	85,82	
	$\sigma$	15,68	

Classificação por  $DIF = O - H$

Dia	Óxido de Nitrogênio (O)	Hidrocarbono (H)	DIF = O-H
7; 9	81; 61	88; 68	-7
3; 5	84; 61	89; 66	-5
1; 8	104; 72	108; 76	-4
2	116	118	-2
3; 10	84; 97	83; 96	1
6	84	81	3
4	77	71	6
$\mu$			-1,14
$\sigma$			4,32

c) *Não* podemos dizer que as duas classes de poluentes estão presentes nas mesmas concentrações pois o coeficiente de variação da variável DIF

$$c_v = \frac{\sigma}{\mu} = -\frac{4,32}{1,14} \approx -379\% \text{ oferece uma expressão de dispersividade alta.}$$

---

8.

9.

10.

11.

12.

13.