2 – REVISÃO DE ESTATÍSTICA DESCRITIVA

Na presente seção faremos uma revisão básica de estatística descritiva, dando ênfase em aspectos operacionais da análise de dados. Como veremos, ao longo dessa seção, analisar os dados e prepará-los para o uso em regressão[1] é fundamental para que se possa obter resultados satisfatórios.

2.1 Tipos de Dados

Existem três tipos básicos de dados no campo da estatística, a saber:

- *Dados Brutos*: 3,4,2,3,2,1,4,2,5,7
- *Rol de Dados*: 1,2,2,2,3,3,4,4,5,7
- *Dados Agrupados em Freqüência*
- *Dados Agrupados em Classes*

Dados agrupados em freqüência

X={1,2,2,2,3,3,4,4,5,7}

X_i	f_i
1	1
2	3
3	2
4	2
5	1
7	1
	10

Dados agrupados em classe

Classe	f_i
0 ├— 2	1
2 ├— 4	5
4 ├— 6	3
6 ├— 8	1
	10

Notação	
a ├— b	$a \le x < b$
a —┤ b	$a < x \le b$
a — b	$a < x < b$
a ├—┤ b	$a \le x \le b$

As freqüências são classificadas em:

- absoluta e relativa
- simples (fi) e acumulada (Fi)

[1] A análise de regressão é o estudo da dependência de uma variável, a variável dependente, em relação a uma ou mais variáveis, variáveis explicativas, com o objetivo de estimar e/ou prever a média da população ou o valor médio da dependente em termos dos valores conhecidos ou fixos (em amostragem repetida) das explicativas. Gujarati, Damodar N. Econometria Básica São Paulo: Makron Books, 2000, pp 21.

Absoluta		
X_i	f_i	F_i
1	1	1
2	3	4
3	2	6
4	2	8
5	1	9
7	1	10
	10	

Relativa		
X_i	f^*_i	F^*_i
1	0,10	0,10
2	0,30	0,40
3	0,20	0,60
4	0,20	0,80
5	0,10	0,90
7	0,10	1,00

ATENÇÃO: Denomina-se amplitude (h) de classe, definida por $\mathbf{h = l_s - l_i}$

onde: l_s = limite superior da classe; e

l_i = limite inferior da classe.

Para efeito de cálculos da variável estudada e desde que não haja impedimento técnico (informações que indiquem, por exemplo, que a classe é fortemente concentrada à direita ou à esquerda) denomina-se $\mathbf{X_i}$ o ponto médio da classe, definido por:

$$X_i = \frac{l_s + l_i}{2}$$

Exemplo:

A tabela abaixo representa os salários pagos a 100 operários de uma empresa. Determinar:

a) freqüência absoluta acumulada, simples relativa e acumulada relativa

b) quantos operários ganham até dois salários mínimos exclusive?

c) quantos operários ganham até 6 salários mínimos exclusive?

d) qual a porcentagem de operários com salários entre 6 e 8 salários mínimos?

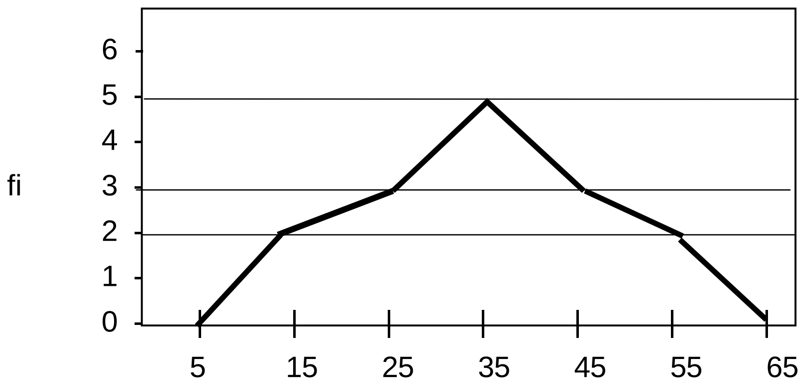
Nº de salários mínimos	Nº de operários (f_i)	F_i	f_i (%)	F_i (%)
0 |--- 2	40			
2 |--- 4	30			
4 |--- 6	10			
6 |--- 8	15			
8 |--- 10	5			
Total	100	-	-	-

2.2. Polígono de Freqüência e Histograma

Considere a seguinte distribuição de freqüência:

X_i	f_i
5	0
15	2
25	3
35	5
45	3
55	2
65	0
	15

Podemos associar a cada valor de X_i o seu correspondente ponto no eixo X e sua freqüência (absoluta ou relativa) ao eixo Y. Teremos então:



Polígono de freqüência
ou
Curva de freqüência

Pode-se fazer uma alteração na distribuição de freqüência da seguinte forma:

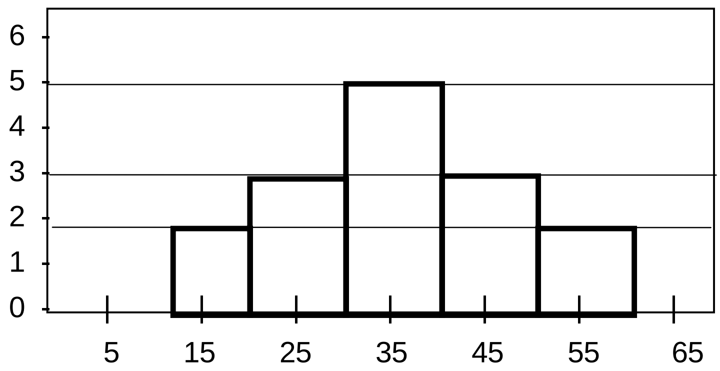
X_i	f_i
0 ├── 10	0
10 ├── 20	2
20 ├── 30	3
30 ├── 40	5
40 ├── 50	3
50 ├── 60	2
60 ├── 70	0
	15

Podemos agora construir um novo desenho, composto por retângulos tais que:

(1) As bases dos retângulos estejam centradas nos pontos médios das classes;

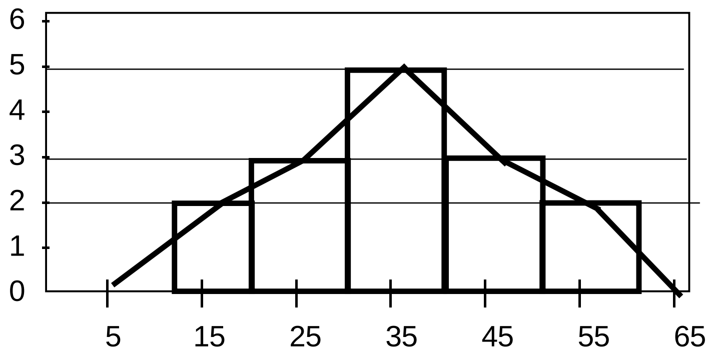
(2) As alturas dos retângulos sejam iguais ou proporcionais às freqüências das respectivas classes.

Teremos então:



[Histograma]

Unindo os dois desenhos, teremos:



NOTA: A área do histograma é igual à área do polígono de freqüência

2.3 Medidas de Posição

As medidas de posição são divididas em dois grandes grupos:

- <u>medidas de tendência central e separatrizes</u> - são valores típicos ou representativo de um conjunto de dados.
- <u>medidas de dispersão</u> - medem o grau de dispersão dos dados, ou seja, o quanto eles se afastam em torno da média.

Cada uma delas se divide em vários tipos, como veremos a seguir.

Medidas de Tendência Central $\left\{\begin{array}{l} \textit{Média} \left\{\begin{array}{l} \textit{Aritmética} \\ \textit{Geométria} \end{array}\right. \\ \\ \textit{Moda} \left\{\begin{array}{l} \textit{Bruta} \\ \textit{Pearson} \\ \textit{Czuber} \end{array}\right. \\ \\ \textit{Mediana} \end{array}\right.$

Separatrizes $\left\{\begin{array}{l} \textit{Mediana} \\ \textit{Decil} \\ \textit{Quartil} \\ \textit{Percentil} \end{array}\right.$

Medidas de Dispersão $\left\{\begin{array}{l} \text{Amplitude Total} \\ \text{Desvio - Médio} \\ \text{Variância} \\ \text{Desvio - Padrão} \\ \text{Coeficiente de Variação} \end{array}\right.$

2.4 Medidas de Tendência Central

2.4.1 Médias

A media é um dos vários indicadores de tendência central que se usa para indicar o ponto na escala de medidas no qual a população (ou amostra) está centrada. É, portanto um valor que indica o centro dos valores de uma coleção de dados (população ou amostra)

A média é o termo médio dos dados na população (ou amostra). Numericamente, é igual a soma das observações dividida pela sua freqüência. É interessante notar que a média é o único valor que, se substituir todos os dados na população (ou amostra), produzirá a mesma soma dos dados originais, e por conseguinte a mesma média.

2.4.1.1 Média Aritmética[2]

Dados $x_1, x_2, x_3, x_4, \ldots, x_n$

$$\boxed{\overline{X} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum x_i}{n}}$$

[2] Mais para frente x será desvio.

Exemplos:

(a) Para dado Bruto: X = 2,3,2,3,5 $\Rightarrow$ $\overline{X} = \frac{2+3+2+3+5}{5} = \frac{15}{5} = 3$

(b) Para dados agrupados em freqüência: $\Rightarrow$ $\overline{X} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{50}{10} = 5$

x_i	f_i	$x_i.f_i$
2	2	4
4	3	12
6	3	18
8	2	16
	10	50

Nesse caso, a média aritmética é chamada de Média Aritmética Ponderada, onde os pesos são os valores f_i .

(c) Para dados agrupados em classe: $\Rightarrow$ $\overline{X} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{46}{10} = 4.6$

classe	f_i	x_i	$x_i.f_i$
0 ⊢ 2	2	1	2
2 ⊢ 4	3	3	9
4 ⊢ 6	2	5	10
6 ⊢ 8	1	7	7
8 ⊢ 10	2	9	18
	10	-	46

IMPORTANTE: Genericamente, a Média Aritmética é expressa por:

$\overline{X} = \frac{\sum x_i \cdot f_i}{\sum f_i}$. Quando $f_i = 1 \Rightarrow$ $\overline{X} = \frac{\sum x_i}{\sum f_i} = \frac{\sum x_i}{n}$

2.4.1.2 Média Geométrica (Mg)

Medida menos afetada pelos valores extremos das observações da série. Muito utilizada em taxas de crescimento.

Dados $x_1, x_2, x_3, x_4, \ldots, x_n$ $\boxed{Mg = \sqrt[n]{x_1.x_2 \ldots x_n}}$

Exemplos:

(a) Para dado Bruto: X = 2,4,8,4 $\Rightarrow$ Mg = $\sqrt[4]{2 \times 4 \times 8 \times 4} = 2^2 = 4$

(b) Para dados agrupados em freqüência:

Nesse caso, as potências de cada valor x_i são as freqüências (f_i) com que eles aparecem da distribuição. Portanto: $Mg = \sqrt[n]{x_1^{f1}.x_2^{f2}....x_n^{fn}}$ e $n = \Sigma f_i$.

x_i	f_i	x_i^{fi}
2	2	2^2
4	3	4^3
6	10	6^{10}
8	5	8^5

Logo: Mg= $\sqrt[20]{2^2 \times 4^3 \times 6^{10} \times 8^5} = 5.25$

(c) Para dados agrupados em classe:

Nesse caso, o valor de x_i utilizado, deverá ser o valor médio da classe

classe	f_i	x_i	x_i^{fi}
10 |— 20	2	15	15^2
20 |— 30	3	25	25^3
30 |— 40	3	35	35^3
40 |— 50	2	45	45^2
	10	-	

Mg = $\sqrt[10]{15^2 \times 25^3 \times 35^3 \times 45^2} = 28.08$

2.4.2 Moda (Mo)

É o valor mais freqüente de uma série de dados.

(a) Para dado bruto:

Dado Bruto	Rol de dados	Moda	Tipo
2,3,4,3,1	1,2,3,3,4	3	unimodal
2,3,2,4,3,5	2,2,3,3,4,5	Mo^1=2; Mo^2=3	bimodal
1,2,2,4,4,3,3	1,2,2,3,3,4,4	Mo^1=2;Mo^2=3;Mo^3=4	trimodal
4,4,2,2,3,3	2,2,3,3,4,4	(ausente)	amodal

(b) Para dados agrupados em freqüência:

x_i	f_i
2	2
4	3
6	4
8	2

Mo =

Tipo =

x_i	f_i
2	1
4	3
6	3
8	2

Mo =

Tipo =

x_i	f_i
10	2
20	2
30	2
40	2

Mo =

Tipo =

(c) Para dados agrupados em classe:

classe	f_i	x_i	F_i
0 |— 20	2	10	
20 |— 40	3	30	
40 |— 60	10	50	
60 |— 80	3	70	
80 |— 100	2	90	
	20	250	

Obs.: Para dados agrupados em classe, existem várias fórmulas de cálculo da Moda. Três, no entanto, são as mais utilizadas.

c.1) Moda Bruta : Ponto Médio da Classe Modal

c.2) Moda de Pearson : $Mo = 3.Md - 2.\overline{X}$

c.3) Moda de Czuber : $Mo = l_i + \left[\dfrac{f_{mo} - f_{ant}}{2f_{mo} - (f_{ant} + f_{post})}\right] \times h$

2.4.3 Mediana (Md)

É o valor que divide o conjunto de dados ordenados em duas metades, com metade dos valores acima da mediana e a metade dos valores abaixo dela. Quando o número de observações (n) é ímpar, a mediana é o valor que ocupa a posição central. Quando n for par, há duas posições centrais no conjunto, então a mediana é a média aritmética dos dois valores que ocupam as posições centrais.

Antes de calcular a mediana, devemos calcular a Posição da Mediana dada por: P(Md)= $\dfrac{n+1}{2}$

(a) Para dado bruto:

Dado Bruto	Rol	P(Md)	Mediana
4,2,3,3,2	2,2,3,3,4	3	3
2,5,4	2,4,5	2	4
2,3,4,2,4	2,2,3,4,4	3	3
5,2,4,4,3,4,2,2	2,2,2,2,4,4,4,5	4,5	3

(b) Para dados agrupados em freqüência:

x_i	f_i	F_i
2	2	2
4	3	5
6	4	9
8	6	15
10	5	20
	20	

Passos para o cálculo da Mediana

(1) Achar n

(2) Calcular F_i

(3) Calcular P(Md)

(4) Procurar P(Md) em F_i

(5) Calcular Md, se for o caso

(c) Para dados agrupados em classe:

classe	f_i	F_i
10 |— 20	2	2
20 |— 30	3	5
30 |— 40	10	15
40 |— 50	3	18
50 |— 60	2	20
	20	

Passos para o cálculo da Mediana

(1) Achar n

(2) Calcular F_i

(3) Calcular P(Md)

(4) Determinar a Classe Mediana

(5) Aplicar a seguinte fórmula: $\boxed{Md = l_i + \dfrac{\frac{n}{2} - F_a}{f_{Md}} \cdot h}$

onde: l_i = limite inferior da classe mediana;

n = número total de observações;

F_a = freqüência acumulada da classe anterior à da mediana

h = amplitude da classe mediana

f_{Md} = freqüência simples da classe mediana.

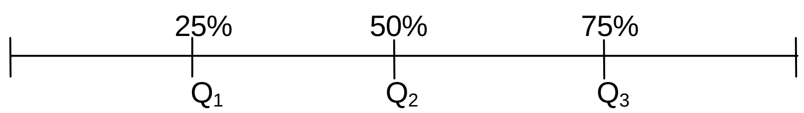
2.5 Separatrizes

2.5.1 A Mediana

É um único valor que divide o rol de dados em duas partes iguais. É importante observar que a mediana é um parâmetro especial de qualquer conjunto de dados, pois é ao mesmo tempo uma medida de tendência central e uma separatriz.

2.5.2 O Quartil

São três valores (Q_1 , Q_2 , Q_3) que dividem o rol (de dados) em quatro partes iguais.



A posição do quartil é dada por:

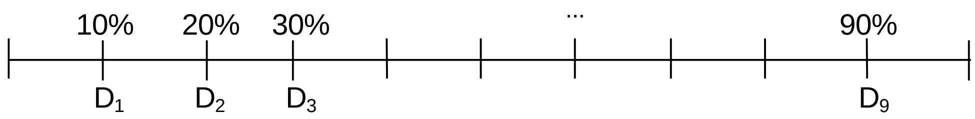
$$\boxed{P(Q_k) = \frac{n \cdot k}{4} + \frac{1}{2}} \quad \text{se n par}$$

$$\boxed{P(Q_k) = \frac{n \cdot k + k}{4}} \quad \text{se n ímpar}$$

k=1,2,3

2.5.3 O Decil

São nove valores que dividem o rol (de dados) em dez partes iguais.

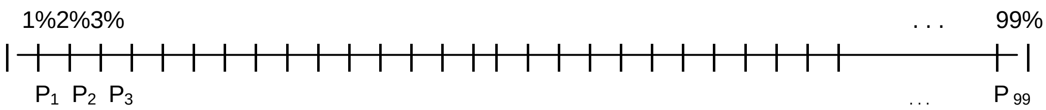


A posição do decil é dada por:

$$\boxed{P(D_k) = \frac{n \cdot k}{10} + \frac{1}{2} \text{ , k=1,2,...,9}}$$

2.5.4 O Percentil

São 99 valores que dividem o rol (de dados) em cem partes iguais.



A posição do percentil é dada por:

$$\boxed{P(P_k) = \frac{n \cdot k}{100} + \frac{1}{2} \text{ , k=1,2,3,...,99}}$$

Exemplos: Calcular a mediana, os três quartis,o 2º e o 7º decis**.**

(a) Rol: { 2 ,2 ,3 ,3 ,4 ,6 ,6 ,6 ,7 ,9 }

$$P(Md) = \frac{n+1}{2} = \frac{10+1}{2} = 5{,}5 \to Md = \frac{4+6}{2} = 5$$

$$P(Q_1) = \frac{n \cdot 1}{4} + \frac{1}{2} = \frac{10 \cdot 1}{4} + \frac{1}{2} = \frac{12}{4} = 3 \to Q_1 = 3$$

$$P(Q_2) = \frac{n \cdot 2}{4} + \frac{1}{2} = \frac{10 \cdot 2}{4} + \frac{1}{2} = \frac{11}{2} = 5{,}5 \to Q_2 = \frac{4+6}{2} = 5$$

$$P(Q3) = \frac{n \cdot 3}{4} + \frac{1}{2} = \frac{10 \cdot 3}{4} + \frac{1}{2} = \frac{16}{2} = 8 \to Q_3 = 6$$

$$P(D_2) = \frac{n \cdot 2}{10} + \frac{1}{2} = \frac{10 \cdot 2}{10} + \frac{1}{2} = 2{,}5 \rightarrow D_2 = \frac{2+3}{2} = 2{,}5$$

$$P(D_7) = \frac{n \cdot 7}{10} + \frac{1}{2} = \frac{10 \cdot 7}{10} + \frac{1}{2} = 7{,}5 \rightarrow D_7 = \frac{6+6}{2} = 6$$

(b) Para dados agrupados em freqüência:

x_i	f_i	F_i
1	5	5
2	8	13
4	27	40
6	30	70
8	20	90
10	10	100
	100	-

$$P(Md) = \frac{n+1}{2} = \frac{100+1}{2} = 50{,}5 \rightarrow Md = 6$$

$$P(Q_1) = \frac{n \cdot 1}{4} + \frac{1}{2} = \frac{100 \cdot 1}{4} + \frac{1}{2} = 25{,}5 \rightarrow Q_1 = 4$$

$$P(Q_2) = \frac{n \cdot 2}{4} + \frac{1}{2} = \frac{100 \cdot 2}{4} + \frac{1}{2} = 50{,}5 \rightarrow Q_2 = 6$$

$$P(Q_3) = \frac{n \cdot 3}{4} + \frac{1}{2} = \frac{100 \cdot 3}{4} + \frac{1}{2} = 75{,}5 \rightarrow Q_3 = 8$$

$$P(D_2) = \frac{n \cdot 2}{10} + \frac{1}{2} = \frac{100 \cdot 2}{10} + \frac{1}{2} = 20{,}5 \rightarrow D_2 = 4$$

$$P(D_7) = \frac{n \cdot 7}{10} + \frac{1}{2} = \frac{100 \cdot 7}{10} + \frac{1}{2} = 70{,}5 \rightarrow D_7 = \frac{6+8}{2} = 7 \text{ [limite de faixa]}$$

CÁLCULO DAS SEPARATRIZES PARA DADOS AGRUPADOS EM CLASSE

$$Mediana\ (Md) = l_i + \frac{\left(\frac{n}{2} - F_a\right) \cdot h}{f_i}$$

$$Quartil\ (Q_k) = l_i + \frac{\left(\frac{n.k}{4} - F_a\right) \cdot h}{f_i}$$

$$Decil\ (D_k) = l_i + \frac{\left(\frac{n.k}{10} - F_a\right) \cdot h}{f_i}$$

$$Percentil\ (P_k) = l_i + \frac{\left(\frac{n.k}{100} - F_a\right) \cdot h}{f_i}$$

onde:
n = número de observações
l_i = limite inferior da classe
F_a = freqüência acumulada da faixa anterior
h = amplitude da classe
f_i = freqüência da classe

Exemplo:

Calcular a mediana, o 3º quartil e o 6º decil :

x_i	f_i	F_i
0 ⊢— 10	8	8
10 ⊢— 20	12	20
20 ⊢— 30	40	60
30 ⊢— 40	20	80
40 ⊢— 50	20	100
	100	-

$$P(M_d) = \frac{100+1}{2} = 50{,}5 \rightarrow Md = 20 + \frac{\left(\frac{100}{2} - 20\right) \cdot 10}{40} = 20 + \frac{(50-20).10}{40} = 27{,}5$$

$$P(Q_3) = \frac{100 \cdot 3}{4} + \frac{1}{2} = 75{,}5 \rightarrow Q_3 = 30 + \frac{\left(\frac{100 \cdot 3}{4} - 60\right).10}{20} = 30 + 7{,}5 = 37{,}5$$

$$P(D_6) = \frac{100 \cdot 6}{10} + \frac{1}{2} = 60{,}5 \rightarrow D_6 = 30 + \frac{\left(\frac{100 \cdot 6}{10} - 60\right).10}{20} = 30 + 0 = 30$$

2.6 Medidas de Dispersão

Existem várias medidas que indicam e medem a dispersão de um conjunto de dados. São elas:

2.6.1 Amplitude Total

É a diferença entre o maior e o menor valor de um conjunto de dados.

$$\boxed{H = AT = MAX - MIN}$$

2.6.2 Desvio Médio

É expresso pela soma dos módulos das diferenças entre os elementos da série e a média, dividida pelo número de elementos da série.

Para dado bruto: $$DM = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$$

Para dado agrupado em freqüência ou em classe: $$DM = \frac{\sum_{i=1}^{n} | x_i - \bar{x} | f_i}{\sum f_i}$$

2.6.3 Variância (VAR ou σ^2)

É a soma dos quadrados das diferenças em relação à média, dividida pelo número de elementos da série.

Para dado bruto: $$\mathrm{VAR} = \frac{\sum (x_i - \bar{x})^2}{n}$$

Para dado agrupado em freqüência ou em classe $$\mathrm{VAR} = \frac{\sum (x_i - \bar{x})^2 \cdot f_i}{\sum f_i}$$

2.6.4 Desvio-Padrão (DP ou σ)

É a raiz quadrada da variância.

Genericamente DP = $\sqrt{\mathrm{VAR}}$

2.6.5 Coeficiente de Variação *(CV)*

É uma medida de dispersão relativa (adimensional) da variação de uma série e é expressa pela razão entre o Desvio-Padrão e a Média.

$$CV = \frac{DP}{\bar{x}} \text{ ou } \%CV = \frac{DP}{\bar{x}} \text{ x } 100$$

NOTA: As duas principais medidas de dispersão são VARIÂNCIA e DESVIO-PADRÃO

Exercícios Propostos

1. Com os dados a seguir, complete as tabelas com as freqüências simples e as acumuladas:

3	2	7	6	9	9	2	9	3	9
6	1	6	5	4	3	3	6	9	7
3	5	10	3	6	9	2	1	9	8
10	10	9	7	5	1	3	1	8	2

X_i	f_i	F_i
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Classe	f_i	F_i
1 |– 3		
3 |– 5		
5 |– 7		
7 |– 9		
9 |– 11		

2. Com os dados da questão anterior, desenhe um histograma com 5 classes e outro com 10 classes.

3. Com os dados a seguir, calcule as medidas de posição: média, moda, mediana, amplitude total, desvio-padrão, variância e coeficiente de variação.

13,47	16,12	14,89	13,50	14,32	15,73	14,27	13,69	14,28	12,47
19,78	17,55	15,00	15,85	17,37	15,30	13,20	14,69	15,22	14,61
13,48	16,51	16,28	14,30	15,35	15,46	11,30	15,87	16,29	16,16
13,74	16,03	15,60	17,68	15,71	14,57	15,34	15,45	12,16	14,23

4. Agrupe os dados do exercício anterior em 6 classes de mesma amplitude e recalcule todas as medidas de posição. Explique porque você encontrou valores diferentes.

5. Os dados a seguir representam a série de lucros mensais de duas empresas distintas (do mesmo setor) para os últimos três anos. Suponha que as duas estão à venda e que você precisa decidir qual delas irá adquirir. Escolha a melhor opção de investimento de acordo com os seus conhecimentos de estatística básica. Você acredita que o seu julgamento seria o mesmo se não conhecesse as medidas de posição?

	ANO 1		ANO 2		ANO 3	
	A	B	A	B	A	B
Jan	12,20	8,51	14,70	14,75	17,66	22,15
Fev	13,11	10,77	15,04	15,59	11,80	7,50
Mar	16,86	20,14	15,00	15,51	14,26	13,66
Abr	15,28	16,20	14,77	14,92	12,73	9,83
Mai	17,59	21,98	14,64	14,59	15,31	16,27
Jun	15,22	16,06	15,35	16,36	11,07	5,66
Jul	11,26	6,16	12,84	10,11	18,97	25,44
Ago	12,58	9,44	12,17	8,43	13,98	12,95
Set	15,88	17,70	14,79	14,98	17,84	22,60
Out	15,76	17,40	14,62	14,56	16,16	18,39
Nov	12,02	8,06	16,21	18,54	15,46	16,64
Dez	17,43	21,56	15,12	15,81	16,84	20,11
ANO	**175,19**	**173,98**	**175,25**	**174,15**	**182,08**	**191,20**

6. Os dados a seguir foram obtidos de duas populações diferentes. Faça uma análise dos valores fornecidos e indique qual o formato que você acredita representar o comportamento de cada uma das populações:

Medida	População I	População J
Média	7,57	6,03
Mediana	6,20	6,20
Moda	5,75	7,00

7. Calcule o 1º e o 3º quartis e o 7º e o 85º percentis das bases de dados a seguir:

2	75	28	30	2	12	68	97	21	8
41	42	32	53	29	79	41	35	14	10
7	67	42	91	91	28	55	15	48	46
47	46	79	28	33	31	67	16	68	7

8. Calcule o 1º e o 3º quartis e o 16º e o 73º percentis de uma distribuição N(5;9)

9. A seguir é fornecida uma base da dados com o salário de 40 analistas de crédito. A primeira metade refere-se a profissionais do sexo feminino e a segunda a profissionais do sexo masculino. Com os instrumentos que você julgar conveniente, avalie se você acredita existir diferenças entre o rendimento por sexo.

2.253,97	2.200,52	2.226,68	2.202,24	2.098,00	2.199,27	2.093,17	2.359,16	2.305,51	2.367,27
2.335,24	2.218,79	2.249,34	2.313,82	2.161,11	2.386,02	2.360,69	2.283,88	2.069,40	2.231,00
2.574,77	2.226,21	2.218,25	2.473,37	2.510,08	2.686,63	2.795,96	2.843,11	2.333,32	2.514,34
3.049,96	2.375,34	2.542,13	2.576,82	2.569,89	2.721,19	2.360,63	2.099,83	3.198,08	2.471,41

10. Abaixo você encontra o tempo de resposta (em dias) entre a aplicação de um tratamento e a efetiva cura de uma certa doença, além de algumas características dos pacientes (levantadas segundo um questionário). Avalie se você consegue determinar algum fator que diferencie, de forma significativa, o tempo de resposta do medicamento à doença.

tempo	sexo	idade	fumante	estado civil
4	masculino	38	não	solteiro
12	masculino	28	não	casado
13	masculino	31	não	viúvo
7	feminino	33	sim	casado
2	masculino	34	sim	solteiro
13	masculino	18	sim	casado
8	masculino	20	sim	solteiro
11	masculino	18	sim	casado
4	feminino	21	não	casado
9	masculino	22	não	casado
11	masculino	19	sim	casado
8	masculino	26	sim	casado
11	feminino	28	sim	casado
17	feminino	29	sim	casado
11	masculino	24	sim	casado
12	feminino	34	não	solteiro
12	feminino	28	sim	casado
12	masculino	30	não	solteiro
9	masculino	26	sim	solteiro
15	feminino	27	sim	casado

*11. Demonstre que a variância da média de duas variáveis aleatórias independentes é igual à média de suas variâncias.

*12. Na estatística podemos encontrar diversas distribuições com um formato previamente conhecido. Dentre elas encontramos a distribuição uniforme. Sabendo que a distribuição uniforme é definida como a representação de pontos com a mesma probabilidade de ocorrência em um determinado intervalo, [a,b], com média = mediana = (a + b) / 2 = e que ela é amodal, indique o formato esperado de um histograma relativo a uma população proveniente de um grupo uniformemente distribuído.

*13. Com os índices (data base jan/94 = 100) relativos às vendas da indústria eletroeletrônica de consumo no Brasil (indicados abaixo), monte uma tabela representativa dos dados agrupados em até sete (7) classes. Com os valores encontrados, monte um histograma e faça uma análise estatística. Utilize seus conhecimentos de medidas de posição para a melhor avaliar a amostra fornecida. Compare os resultados com o seu conhecimento a respeito da evolução deste segmento durante o período indicado.

Data	Índice
31/01/98	180,02
28/02/98	192,45
31/03/98	261,97
30/04/98	247,51
31/05/98	227,16
30/06/98	194,29
31/07/98	196,33
31/08/98	223,44
30/09/98	212,61
31/10/98	249,55
30/11/98	298,34
31/12/98	248,48
31/01/99	151,11
28/02/99	101,98
31/03/99	185,44
30/04/99	176,49
31/05/99	140,98
30/06/99	178,36
31/07/99	191,89
31/08/99	204,02
30/09/99	190,90
31/10/99	223,93
30/11/99	259,44

Fonte: Eletros.

*14. Quais as vantagens e desvantagens você obtém ao utilizar, numa análise estatística, uma tabela com dados agregados?