Respostas sugeridas para a Lista 1

1 Um artigo reportou dados sobre um experimento, investigando o efeito de muitas variáveis de processos de oxidação, em fase vapor, de naftaleno. Uma amostra de conversão percentual molar de naftaleno em anidrido maléico resulta em:

4, 2 ; 4, 7 ; 4, 7 ; 5, 0 ; 3, 8 ; 3, 6 ; 3, 0 ; 5, 1 ; 3, 1 ; 3, 8 ; 4, 8 ; 4, 0 ; 5, 2 ; 4, 3 ; 2, 8 ; 2, 0 ; 2, 8 ; 3, 3 ; 4, 8 ; 5, 0 ; 4, 8 ; 3, 9 ; 5, 3 ; 5, 0 ; 4, 7 ; 3, 6 ; 3, 8 ; 3, 0 ; 3, 2 ; 4, 2 ; 4, 5 ; 4, 7 ; 4, 9 ; 4, 0 ; 4, 1 ; 4, 4 ; 5, 0

a) Encontre a média, a mediana e a moda;

Temos $\sum x_i = 4,2 + 4,7 + 4,7 + 5,0 + 3,8 + 3,6 + 3,0 + 5,1 + 3,1 + 3,8 + 4,8 + 4,0 + 5,2 + 4,3 + 2,8 + 2,0 + 2,8 + 3,3 + 4,8 + 5,0 + 4,8 + 3,9 + 5,3 + 5,0 + 4,7 + 3,6 + 3,8 + 3,0 + 3,2 + 4,2 + 4,5 + 4,7 + 4,9 + 4,0 + 4,1 + 4,4 + 5,0 = 153,1$, ou seja,

$$\overline{x} = \frac{153,1}{37} \approx 4,1378.$$

Para a mediana, primeiro colocamos os dados em ordem (rol):

2,0; 2,8; 2,8; 3,0; 3,0; 3,1; 3,2; 3,3; 3,6; 3,6; 3,8; 3,8; 3,8; 3,9; 4,0; 4,0; 4,1; 4,2; 4,2; 4,3; 4,4; 4,5; 4,7; 4,7; 4,7; 4,7; 4,8; 4,8; 4,8; 4,9; 5,0; 5,0; 5,0; 5,0; 5,1; 5,2; 5,3

Como temos 37 valores e este é um número ímpar, a mediana é o termo central na posição $\frac{37+1}{2} = 19$, ou seja, 4,2 e, para a moda, é fácil ver que 4,7 e 5,0 apareceram ambos 4 vezes.

c) Organize os dados em uma tabela de frequências adequada.

1° passo: A = maior valor − menor valor $= 5,3 - 2,0 = 3,3$.

2° passo: $k = \sqrt{n} = \sqrt{37} \approx 6$.

3° passo: $c = \frac{A}{k-1} = \frac{3,3}{5} = 0,66$.

4° passo: LI_1 = menor valor $- \frac{c}{2} = 2,0 - \frac{0,66}{2} = 1,67$.

Assim, temos

naftaleno	f_i	f_r	$f_p(\%)$
$1,67 \vdash 2,33$	1	0,02702	2,702
$2,33 \vdash 2,99$	2	0,05405	5,405
$2,99 \vdash 3,65$	7	0,18919	18,919
$3,65 \vdash 4,31$	10	0,27027	27,027
$4,31 \vdash 4,97$	10	0,27027	27,027
$4,97 \vdash 5,63$	7	0,18919	18,919
Total	37	1,00	100

2 Acredita-se que a resistência à tensão da borracha siliconizada seja uma função da temperatura de cura. Um estudo foi realizado, no qual amostras de 12 espécimes de borracha foram preparadas usando temperaturas de cura de 20° C e 45° C. Os dados mostram os valores de resistência à tensão, em megapascals:

a) Identifique a variável em estudo e classique-a;

resistências medidas para uma dada temperatura, ou seja, uma variável quantitativa contínua, por ser oriunda de uma medição.

b) Faça uma análise descritiva comparando os dois grupos e interpretando os resultados a partir de média, mediana e uma medida adequada, em termos de variabilidade;

Tomando o desvio padrão como medida de variabilidade, temos

temperatura	$\overline{x}$	Md	σ
$20°\ C$	2,1075	2,1	0,06783
$45°\ C$	2,235	2,2	0.19452

onde vemos que todas as medidas aumentaram conforme a temperatura, em especial, o desvio padrão triplicou indicando perda de uniformidade nas medições.

3 Uma equipe de Higiene e Segurança do Trabalho de uma empresa de aviação, preocupada com o número de horas trabalhadas pelos funcionários, observou em seus registros recentes, uma amostra com o tempo de mão-de-obra gasto na revisão completa de um motor de jato. A tabela obtida foi:

Tempo (horas)	f_i
$0,00 \vdash 4,00$	1
$4,00 \vdash 8,00$	5
$8,00 \vdash 12,00$	10
$12,00 \vdash 16,00$	15
$16,00 \vdash 20,00$	4

a) Determine o número médio de horas de mão de obra necessário para revisão de cada motor;

Para as 5 classes de frequência, os pontos médios P_i são 2, 6, 10, 14, 18. Assim, lembrando que $n = \sum f_i$, temos

$$\begin{aligned}\overline{x} &= \frac{\sum_{i=1}^{5} P_i f_i}{n} \\ &= \frac{2\cdot 1 + 6\cdot 5 + 10\cdot 10 + 14\cdot 15 + 18\cdot 4}{35} \\ &= \frac{2 + 30 + 100 + 210 + 72}{35} \\ &= \frac{414}{35} \\ &\approx 11,8286\end{aligned}$$

b) Com base nesta informação (item a)), qual deve ser o tempo total de mão de obra para a revisão de dez motores que aguardam revisão? $10 \cdot 11,83 = 118,3$ horas

c) Se a empresa dispõe no momento de dois homens trabalhando 12 horas por dia nestas revisões, conseguirá provavelmente revisar estes dez motores em quatro dias? Por que? Não. Com a média calculada, sabemos que cada homem fará um motor por dia de modo que precisarão de pelo menos 5 dias para os dez motores.

4 Foram anotados os níveis de colesterol (em mg/100ml) para uma amostra de trinta pacientes de uma clínica cardíaca. As medidas se referem a homens entre 40 e 60 anos de idade que foram à clínica fazer um *check-up*.

a) Indique a variável em estudo e classifique-a; Variável: níveis de colesterol (em mg/100ml), que é quantitativa contínua (por vir de uma medição) mas nesse caso, como os valores anotados são inteiros, podemos considerá-la como quantitativa discreta.

b) Calcule o nível de colesterol mediano.

Temos $\sum x_i = 160+161+160+170+167+163+172+172+173+177+178+182+181+181+186+185+194+197+199+203+205+203+206+206+211+209+208+214+218+225 = 5666$, ou seja,

$$\overline{x} = \frac{5666}{30} \approx 188,8666.$$

c) Organize os dados em uma tabela de frequência completa com intervalos de classes de tamanho 10, iniciando do nível 160;

1° passo: A = maior valor − menor valor = 225 − 160 = 65.

2°, 3° e 4° passo: Como vamos usar classes de tamanho 10, temos $k = 7$ para $c = 10$ e também já foi dado $LI_1 = 160$.

Assim, temos

colesterol	f_i	f_r	$f_p(\%)$
$160 \vdash 170$	5	0,16667	16,667
$170 \vdash 180$	6	0,20000	20,000
$180 \vdash 190$	5	0,16667	16,667
$190 \vdash 200$	3	0,10000	10,000
$200 \vdash 210$	7	0,23333	23,333
$210 \vdash 220$	3	0,10000	10,000
$220 \vdash 230$	1	0,03333	3,333
Total	37	1,00	100

d) Refaça o item b) usando as informações da tabela de frequencias obtida em c) e comente as diferenças encontradas entre os valores das medidas calculadas em b) e c).

Para as 7 classes de frequência, os pontos médios P_i são 165, 175, 185, 195, 205, 215, 225. Assim, temos

$$\begin{aligned}\overline{x} &= \frac{\sum_{i=1}^{7} P_i f_i}{n} \\ &= \frac{165\cdot 5 + 175\cdot 6 + 185\cdot 5 + 195\cdot 3 + 205\cdot 7 + 215\cdot 3 + 225\cdot 1}{30} \\ &= \frac{5690}{30} \\ &\approx 189,6666\end{aligned}$$

Note a pequena diferença obtida entre as médias que pode ser entendida por tomarmos a média da tabela como a média dos pontos médios sendo *ponderada* pelas frequências. Se usássemos outra medida com uma proporção melhor das frequências, ou um número maior de classes (mais pesos) talvez obteríamos um valor mais próximo entre as médias.

5

6

7

8 A média e o desvio padrão da produtividade de duas cultivares de milho são respectivamente $\overline{x}_A = 4,0$ t/ha e $s_A = 0,80$ t/ha para a variedade de polinização aberta A e $\overline{x}_B = 8,0$ t/ha e $s_B = 1,20$ t/ha para o híbrido simples B. Qual das cultivares possui maior uniformidade de produção?

Por "maior uniformidade", entenda "menor variabilidade", ou seja, menor dispersão entre os dados. Como nos é dada a média e o desvio padrão, a medida de variabilidade que os usa diretamente é o coeficiente de variação CV, e daí, temos

$$CV_A = \frac{s_A}{\overline{x}_A}\cdot 100 = \frac{0,8}{4}\cdot 100 = 20 \quad \text{e} \quad CV_B = \frac{1,2}{8}\cdot 100 = 15.$$

Portanto, o cultivar B é o mais uniforme.

9

10 Uma distribuidora de refrigerantes fez um levantamento sobre o consumo semanal (em litros) por pessoa, em jan/2005, em uma cidade do litoral, obtendo a tabela abaixo:

Consumo	n° de pessoas
$0,0 \vdash 0,5$	10
$0,5 \vdash 1,0$	25
$1,0 \vdash 1,5$	9
$1,5 \vdash 2,0$	7
$2,0 \vdash 2,5$	6

a) Determine e interprete o consumo médio.

Para as 5 classes de frequência, os pontos médios P_i são 0,25; 0,75; 1,25; 1,75 e 2,25. Assim, temos

$$\begin{aligned}\overline{x} &= \frac{\sum_{i=1}^{5} P_i f_i}{n} \\ &= \frac{0,25\cdot 10 + 0,75\cdot 25 + 1,25\cdot 9 + 1,75\cdot 7 + 2,25\cdot 6}{57} \\ &= \frac{58,25}{57} \approx 1,02193\end{aligned}$$

b) Qual o percentual de pessoas que consomem menos de 1 litro por semana?

Basta ver que as pessoas que consomem menos de 1 litro por semana correspondem à primeira e segunda classe, ou seja, 35. Assim, $\frac{35}{57} \approx 0,614 = 61,4\%$.

c) Determine e interprete o consumo modal e o consumo mediano.

Temos LI_{Mo} é o limite inferior da classe modal, aquela que possui maior frequência simples, ou seja, a segunda que é 0,5.

Δ_1 é a diferença entre as frequências simples da classe modal e a classe anterior, ou seja, $25 - 10 = 15$;

Δ_2 é a diferença entre as frequências simples da classe modal e a classe posterior, $25 - 9 = 16$;

c_{Mo} é a amplitude da classe modal, que é 0,5. Daí,

$$\begin{aligned}Mo &= LI_{Mo} + \left(\frac{\Delta_1}{\Delta_1+\Delta_2}\right)\cdot c_{Mo} \\ &= 0,5 + \left(\frac{15}{15+16}\right)\cdot 0,5 \\ &= 0,5 + (0,483870968)\cdot 0,5 \\ &= 0,5 + 0,241935484 = \boxed{0,741935484}\end{aligned}$$

Repare como a moda é próxima do ponto médio da classe modal.

Para a mediana, o primeiro passo aqui é identificar a classe dela. Lembre-se que a mediana é uma posição no conjunto de dados em rol, mas especificamente na posição $\frac{n+1}{2}$, que nosso caso é $\frac{57+1}{2} = 29$. Então, a mediana está na posição 29 e nossos dados já estão organizados na tabela, assim precisamos identificar em qual classe está essa posição. Repare que, pelas frequências, os primeiros 10 valores estão na primeira classe e os 25 seguintes estão na segunda, o que nos permite assumir que esta é a classe da mediana ($29° < 35°$ valor que delimita a segunda classe). Logo:

LI_{Md} é o limite inferior da classe mediana, no caso, $0,5$;

f_{Md} é a frequência simples da classe mediana, que é 25;

F_A é a frequência acumulada das classes anteriores à classe mediana, no caso, temos apenas uma classe anterior com frequência 10;

c_{Md} é a amplitude da classe mediana, no caso, é $0,5$.

Assim, temos

$$\begin{aligned} Md &= LI_{Md} + \left(\frac{\frac{n}{2} - F_A}{f_{Md}}\right) \cdot c_{Md} \\ &= 0,5 + \left(\frac{\frac{57}{2} - 10}{25}\right) \cdot 0,5 \\ &= 0,5 + \left(\frac{18,5}{25}\right) \cdot 0,5 \\ &= 0,5 + (0,74) \cdot 0,5 \\ &= 0,5 + 0,37 = \boxed{0,87} \end{aligned}$$

d) Se a empresa tem um lucro de R$0,50 por litro, qual o lucro médio por pessoa?

Pelo consumo médio já calculado, podemos assumir que o lucro médio será $0,5 \cdot 1,02193 = 0,510965 \approx R\$0,511$.

11 Um órgão do governo do estado está interessado em determinar padrões sobre o investimento em educação, por habitante, realizado pelas prefeituras. De um levantamento amostral com 10 cidades, foram obtidos os valores da tabela abaixo:

Cidade	A	B	C	D	E	F	G	H	I	J
Invest.	20	16	14	8	19	15	14	16	19	18

Nesse caso, será considerado como investimento básico a média final das observações calculada da seguinte maneira:

1. Obtém-se uma média inicial.

2. Eliminam-se do conjunto aquelas observações que forem superiores à média inicial mais duas vezes o desvio padrão, ou inferiores à média inicial menos duas vezes o desvio padrão. (Este procedimento tem a finalidade de eliminar do conjunto a cidade cujo investimento é muito diferente dos demais)

3. Calcula-se a média final com o novo conjunto de observações.

Qual o investimento básico que você daria como resposta?

Precisamos estabelecer a média inicial:

$$\begin{aligned} \bar{x} &= \frac{\sum x_i}{n} \\ &= \frac{20+16+14+8+19+15+14+16+19+18}{10} \\ &= \frac{159}{10} = 15,9 \end{aligned}$$

e agora, calculamos o desvio padrão:

$$\begin{aligned} \sigma &= \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \\ &= \sqrt{\frac{(20-15,9)^2 + \ldots + (18-15,9)^2}{10}} \\ &= \sqrt{\frac{110,9}{10}} \\ &= \sqrt{11,09} \\ &\approx 3,33 \end{aligned}$$

Agora, no passo 2 eliminamos os valores que estiverem fora do intervalo $[\bar{x} - 2\sigma; \bar{x} + 2\sigma] = [9,24; 22,56]$. Repare que nenhum valor passa de 20 mas temos um valor abaixo da margem de $9,24$, no caso, o valor 8.

Finalmente, recalculamos a média:

$$\begin{aligned} \bar{x} &= \frac{\sum x_i}{n} \\ &= \frac{20+16+14+19+15+14+16+19+18}{9} \\ &= \frac{151}{9} \approx \boxed{16,7777} \end{aligned}$$

Portanto, o investimento básico será $16,7777$.

12

13 Foi realizada na região Oeste do Paraná, no município de Marechal Cândido Rondon, em 1992, um levantamento da produtividade leiteira diária de 20 produtores rurais, atendidos pelo plano "Panela Cheia"(Roesler, 1997). Os resultados dos intervalos de parto (em meses) dos 20 produtores estão apresentados a seguir.

11,80	14,55	11,90	14,65	12,00	14,70	12,30	15,00
12,80	15,10	12,99	15,20	13,10	15,50	13,50	
15,80	13,80	15,90	14,10	15,96			

Obtenha as seguintes estimativas das medidas de dispersão:

a) Amplitude total; $15,96 - 11,80 = 4,16$

b) Variância e desvio padrão; $\sigma^2 = 1,87477875$ e $\sigma = \sqrt{\sigma^2} \approx 1,3692256$.

c) Coeficiente de variação; $CV = \frac{\sigma}{\mu} = \frac{1,36922560}{14,0325} = 0,097575314$

d) Em cada caso anterior comentar, sobre o significado da estimativa obtida e sobre a forma que devem ser aplicadas;

Revise as definições e usos de cada uma...

e) Com a relação à Curtose, qual é a classificação destes dados?

Não pedida mas se desejar, revise a fórmula da Curtose e encontrará $K = -1,06$. Interpretação: Curtose baixa (platicúrtica).

f) Se cada dado for dividido por 12, para se obter o intervalo de partos em anos, qual será os novos valores da amplitude, variância, desvio padrão, CV e erro padrão da média?

Só refazer as contas...

g) Se você fosse solicitado a representar os dados por duas medidas, quais você usaria e por que?

Resposta pessoal...

h) Após o programa Panela Cheia o intervalo de partos apresentou média de 13, 85 e desvio padrão de 2, 00 meses. Qual é na sua opinião a situação que apresentou maior variabilidade, antes ou após o Programa?

Com os referidos dados, o *CV* após o programa foi para 0, 144404332, ou seja, a variabilidade aumentou depois do programa.

14 Abaixo estão representados os dados referentes a um grupo de animais avaliados pela idade, sexo, espécie, e nível de infestação por protozoários.

Animal	Idade	Sexo	Espécie	Nível de Infestação
1	2	Macho	Bovino	Alta
2	2	Macho	Suíno	Baixa
3	1	Fêmea	Caprino	Média
4	4	Fêmea	Bovino	Média
5	5	Fêmea	Caprino	Alta
6	2	Fêmea	Caprino	Alta
7	4	Macho	Suíno	Média
8	2	Fêmea	Suíno	Média
9	5	Macho	Bovino	Baixa
10	5	Fêmea	Bovino	Média
11	3	Macho	Caprino	Alta
12	3	Macho	Caprino	Média
13	5	Macho	Bovino	Média
14	1	Fêmea	Suíno	Baixa

a) Identifique e classifique todas as variáveis descritas no banco de dados acima; Fácil...

b) Construa tabelas de distribuição de frequência para as variáveis idade e sexo;

Idade	n° de animais
1	2
2	4
3	2
4	2
5	4

Sexo	n° de animais
Macho	7
Fêmea	7

c) Construa uma tabela de dupla entrada (ou tabela de contingência) para as variáveis espécie e nível de infestação e faça uma análise da tabela;

	Nível de infestação		
Espécie	Alta	Média	Baixa
Bovino	1	2	1
Caprino	3	2	0
Suíno	0	2	2

d) Represente com gráfico adequado as variáveis espécie e nível de infestação.

Aproveitei a tabela de contingência e fiz um gráfico já com as duas variáveis mas você podia seguir o modelo de exemplo do livro ou fazer gráficos individuais para cada uma.