

Atividade de Probabilidade e Estatística para resolução com linguagem R

Definições:

* Análise Exploratória de Dados

: - O objetivo último da Estatística é tirar conclusões sobre o todo a partir de informações fornecidas por parte representativa do todo. Assim, realizadas as fases de Estatística Descritiva, fazemos uma análise dos resultados obtidos, através dos métodos da Estatística Indutiva ou Inferencial, que tem por base a indução ou inferência, e tiramos desses resultados, conclusões e previsões.

* Tipos de Variáveis

:a - Qualitativa - quando seus valores são expressos por atributos: sexo, cor da pele, etc.

:b - Quantitativa - quando seus valores são expressos em números. Uma variável quantitativa que pode assumir, teoricamente, qualquer valor entre dois limites recebe o nome de variável contínua; uma variável que só pode assumir valores pertencentes a um conjunto enumerável recebe o nome de variável discreta.

* Distribuição de Frequência Intervalar

: - Ao agrupar valores de uma variável em intervalos, poderá ser feita uma Distribuição de Frequência Intervalar, de forma a ganharmos em simplicidade, mas perdermos em pormenores, serve basicamente para simplificar a análise dos dados.

* Distribuição de Frequência Nominal

: - Quando as variáveis a analisar são qualitativas.

* Distribuição de Frequência por Valor

: - Quando as variáveis a analisar são quantitativas.

* Medidas de Posição

- Média

: - Média aritmética é o quociente da divisão da soma dos valores da variável pelo número deles: $\{M = Sx/n\}$, sendo "M" a média aritmética, "x" os valores da variável, "n" o número de valores e "S" representando a soma.

- Mediana

: - A mediana é outra medida de posição definida como o número que se encontra no centro de uma série de números, estando estes dispostos segundo uma ordem. Em outras palavras, a mediana de um conjunto de valores, ordenados segundo uma ordem de grandeza, é o valor situado de tal forma no conjunto que o separa em dois subconjuntos de mesmo número de elementos.

- Moda

: - Denominamos moda o valor que ocorre com maior frequência em uma série de valores.

* Medidas de Variabilidade

- Variância

: - A variância baseia-se nos desvios em torno da média aritmética, porém determinado a média aritmética dos quadrados dos desvios. Assim, representando a variância por $\{var = M(x^2) - M(x)^2\}$, sendo "var" a variância, "M" a média aritmética e "x" os valores da variável.

- Desvio Padrão

: - Sendo a variância calculada a partir dos quadrados dos desvios, ela é um número em unidade quadrada em relação à variável em questão, o que, sob o ponto de vista prático, é um inconveniente. Por esse motivo, imaginou-se uma nova medida que tem utilidade e interpretação práticas, denominada desvio padrão, definida como a raiz quadrada da variância $\{dp = (M(x^2) - M(x)^2)^{1/2}\}$, sendo "dp" o desvio padrão, "M" a média aritmética e "x" os valores da variável.

* Box Plot

: - O boxplot é um gráfico utilizado para avaliar a distribuição empírica dos dados. O boxplot é formado pelo primeiro e terceiro quartil e pela mediana. As hastes inferiores e superiores se estendem, respectivamente, do quartil inferior até o menor valor não inferior ao limite e do quartil superior até o maior valor não superior ao limite superior. Os limites são calculados da forma abaixo: Limite inferior - $\max\{\min(\text{dados}); Q1 - 1,5(Q3 - Q1)\}$. Limite superior - $\min\{\max(\text{dados}); Q3 + 1,5(Q3 - Q1)\}$.

Questões:

1) Os dados a seguir referem-se ao número de livros adquiridos, no ano passado, pelos 40 alunos da Turma A:

4	2	1	0	3	1	2	0	2	1
0	2	1	1	0	4	3	2	3	5
8	0	1	6	5	3	2	1	6	4
3	4	3	2	1	0	2	1	0	3

- a) Organize os dados em uma distribuição de frequência por valor.
- b) Qual o percentual de alunos que adquiriram menos do que 3 livros?
- c) Qual o percentual de alunos que adquiriram pelo menos 4 livros?

2) Considere os dados abaixo referentes ao consumo em água, em m³, de 75 contas da CORSAN:

32	06	22	11	34	40	16	26	23	31	27	10	38	17	13
45	25	50	18	23	35	22	30	14	18	20	13	24	35	29
33	48	20	12	31	39	17	58	19	16	12	21	15	12	20
51	12	19	15	41	29	25	13	23	32	14	27	43	37	21
28	37	26	44	11	53	38	46	17	36	28	49	56	19	11

- a) Organize os dados numa distribuição de frequência intervalar com 9 classes de amplitudes iguais.
- b) Faça um histograma

3) Para se estudar o desempenho de duas corretoras de ações, selecionou-se de cada uma delas, amostras aleatórias das ações negociadas. Para cada ação selecionada computou-se a porcentagem de lucro apresentada durante um período fixado de tempo. Os dados são os seguintes:

Corretora A	Corretora B
45 60 54	57 55 58
62 55 79	50 52 59
38 48 64	59 55 56
55 56 55	61 52 53
54 59 48	57 57 50
65 55 60	55 58 54
	56 51 56

Faça uma Análise Exploratória de dados e verifique qual corretora apresenta o melhor desempenho. Utilize todos os conceitos pesquisados para chegar a uma conclusão!

Obs.: Para diminuir o trabalho, os dados das matrizes apresentadas nas questões estão nos arquivos numerados de acordo com estas, bom desafio!

Para dúvidas ou esclarecimentos, enviar e-mail para: michael.cstrc@gmail.com