

Exercícios de Estatística e Estrutura de Dados com Python

1) A tabela abaixo se refere à dureza de 30 peças de alumínio retiradas como uma amostra para avaliar um processo de fabricação. A respeito desses dados, responda os itens a seguir:

- Apresente o boxplot dessa amostra e identifique se existe algum “outlier” (apenas apresente os valores desses outliers). Faça com que seu programa apresenta automaticamente os outliers.
- Nesse processo, qual é o valor de dureza onde estão 95% dos menores valores de dureza?
- Nesse processo, qual seria a faixa de valores, onde encontráramos 80% dos valores de dureza?
- Após essa primeira análise, implemente um programa que realize a remoção automática dos outliers dos dados e, em seguida, apresente um novo boxplot e as novas medidas de média e desvio-padrão.

53.0	70.2	84.3	69.5	77.8	87.5
53.4	82.5	67.3	54.1	70.5	71.4
95.4	51.1	74.4	55.7	63.5	85.8
53.5	64.3	82.7	78.5	55.7	69.1
72.3	59.5	55.3	73.0	52.4	50.7

2) Em um estudo sobre lesões causadas durante a prática de esportes, 25 escolas do Brasil foram selecionadas e entrevistadas. Foram coletados os dados abaixo, sobre o número de lesões graves em atletas do sexo masculino em duas modalidades de esporte. A respeito desses dados, responda os itens abaixo:

Basquete:	1	2	4	4	7
	3	3	2	4	5
	2	4	3	5	3
	2	4	3	6	5
	5	6	4	6	5
Futebol:	1	7	7	6	1
	2	6	1	7	2
	1	3	2	7	5
	6	1	7	4	1
	5	7	6	3	2

- Armazene os dados apresentados acima na forma de um DICT, para facilitar as operações seguintes.
- Uma segunda rodada de coleta de dados foi realizada. Nesse caso, 3 novas escolas foram adicionadas aos dados, cujos valores de lesões são os seguintes:
 - Basquete: [0, 10, 5]
 - Futebol: [2, 15, 4]

Adicione esses novos valores aos DICT criado no item (a). A partir desse ponto, todas as análises estatísticas devem ser feitas com esses novos dados.

- c) Determine a média, a mediana e o desvio-padrão de cada modalidade de esporte.
- d) Identifique e explique se os dados indicam se existe uma maior chance de lesões graves em uma modalidade de esporte. Aqui a explicação pode ser dada utilizando-se um boxplot e os dados de tendência central (média e mediana)
- e) Identifique se existem “outliers” nos dados, faça a sua remoção automática e responda novamente o item (d).
- 3) Os dados abaixo apresentam as notas de 18 alunas do 1º ano de faculdade e também de 20 alunos do 1º ano. A partir dessa amostra, responda os itens a seguir:
- Determine a média e a mediana das notas dos alunos e das alunas. Por que temos uma média maior do que a mediana para ambas as distribuições?
 - A distribuição das alunas apresenta um “outlier” alto. Remova esse “outlier” e recalcule a média e a mediana de ambas as notas.
 - Apresente o boxplot comparando as notas dos alunos, das alunas e de toda a turma. A partir dessa visualização, podemos concluir que existe uma diferença de desempenho entre cada grupo do estudo?

Alunas	154	109	137	115	152	140	154	178	101
	103	126	126	137	165	165	129	200	148
Alunos	108	140	114	91	180	115	126	92	169
	109	132	75	88	113	151	70	115	146
									104

- 4) A tabela de dados abaixo contém informações de 30 funcionários de uma empresa. Nessa tabela, temos a informação de salário, anos de experiência, educação e se o funcionário possui ou não um cargo de gestão. A partir dessa tabela, faça as análises a seguir:
- Monte os boxplots adequados de forma que você possa analisar se existe uma relação entre o salário e a formação do funcionário (independentemente se ele possui um cargo de gestão ou não)
 - Monte os boxplots adequados de forma a analisar a importância da educação no salário do funcionário, porém agora faça uma distinção se o funcionário possui ou não um cargo de gestão.

Salário	Experiência (anos)	Educação	Gestão
13876	1	Bacharel	Sim
11608	1	Doutorado	Não
18701	1	Doutorado	Sim
11283	1	Mestrado	Não
11767	1	Doutorado	Não
20872	2	Mestrado	Sim
11772	2	Mestrado	Não
10535	2	Bacharel	Não
12195	2	Doutorado	Não
12313	3	Mestrado	Não
14975	3	Bacharel	Sim
21371	3	Mestrado	Sim
19800	3	Doutorado	Sim
11417	4	Bacharel	Não
20263	4	Doutorado	Sim
13231	4	Doutorado	Não

Salário	Experiência (anos)	Educação	Gestão
12884	4	Mestrado	Não
13245	5	Mestrado	Não
15965	5	Bacharel	Sim
12336	6	Bacharel	Não
21352	6	Doutorado	Sim
13839	6	Mestrado	Não
16978	7	Bacharel	Sim
14803	8	Mestrado	Não
22184	8	Doutorado	Sim
13548	8	Bacharel	Não
14467	10	Bacharel	Não
15942	10	Mestrado	Não
23174	10	Doutorado	Sim
23780	10	Mestrado	Sim

5) Com o objetivo de melhorar o nível de seus funcionários em cargos de chefia, uma empresa montou um treinamento (formado por três disciplinas: direito, política e estatística) e indicou 25 funcionários para a primeira turma. Os dados referentes à seção a que pertencem esses funcionários e as suas notas em cada disciplina estão apresentados na tabela abaixo. Em relação à esses dados, responda os itens a seguir:

- a) Calcule a média, a mediana e o desvio padrão da turma para cada disciplina.
- b) A partir da construção dos boxplots de cada disciplina, compare e indique as diferenças existentes entre os resultados dos funcionários.
- c) Crie uma última coluna na tabela que contém a média das notas de cada um dos 25 funcionários. Em seguida, apresente o boxplot dessas médias.
- d) A partir dos dados da coluna criada no item anterior, compare os boxplots da média dos funcionários por seção. Ou seja, crie um gráfico com 4 boxplots (pessoal, técnica, vendas e turma toda) e explique se podemos identificar alguma diferença de desempenho entre cada seção da empresa.

Seção	Direito	Política	Estatística
Pessoal	9	9,0	9
Pessoal	9	6,5	9
Pessoal	9	9,0	8
Pessoal	9	6,0	8
Pessoal	9	6,5	9
Pessoal	9	6,5	10
Pessoal	9	9,0	8
Técnica	9	6,0	8
Técnica	9	9,0	9
Técnica	9	9,0	8
Técnica	9	7,0	10
Técnica	9	5,5	7
Técnica	9	6,0	7
Técnica	9	8,0	9
Vendas	9	7,0	8
Vendas	9	9,0	7
Vendas	9	10,0	8
Vendas	9	5,5	9
Vendas	9	7,0	2
Vendas	9	6,0	7
Vendas	9	6,5	7
Vendas	9	6,0	8
Vendas	9	9,0	9
Vendas	9	6,5	8
Vendas	9	7,0	7