

Para cada exercício a seguir, utilize a linguagem Python para os cálculos e operações solicitadas.

Exercício 1 (IME/USP - <http://www.ime.usp/giapaula/cursosgrad.htm>)

A tabela abaixo mostra o numero de meses em que houve aumento do nível de atividade de quinze empresas de tamanho pequeno (P), médio (M) e grande (G), do setor comercial (C) e industrial (I).

| Empresa | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|
| Meses | 8 | 9 | 4 | 5 | 3 | 6 | 8 | 6 | 6 | 8 | 5 | 5 | 6 | 4 | 4 |
| Setor | C | C | I | I | I | C | C | I | I | C | C | I | C | I | I |
| Tamanho | G | M | G | M | M | P | G | M | P | M | P | P | M | M | G |

- Divida as empresas em dois grupos: comércio (C) e indústria (I). Compare os grupos em relação à média e à mediana do número de meses com crescimento.
- Calcule o desvio padrão e a variância para os dois grupos. Qual dos grupos é mais homogêneo em relação ao número de meses com crescimento?
- Calcule a média, mediana, desvio padrão e variância do número de meses com crescimento para os três tamanhos (P, M e G). Compare essas medidas. Com base nessa análise, você diria que existe relação entre o tamanho da empresa e o número de meses com crescimento?

Exercício 2 (IME/USP - <http://www.ime.usp/giapaula/cursosgrad.htm>)

O peso (em Kg) de 30 mulheres com 168 cm de altura, segundo a idade (em anos) é apresentado abaixo:

| Idade | Peso | | | | |
|-------|------|--|----|----|----|
| 40 | 55 | | 50 | 68 | 65 |
| 45 | 58 | | 56 | 62 | 65 |
| 50 | 60 | | 74 | 70 | 78 |
| 55 | 77 | | 78 | 70 | 72 |
| 60 | 70 | | 76 | 74 | 83 |
| 65 | 65 | | 82 | 72 | 82 |

- Calcule a média, mediana, desvio padrão e variância para o peso dos seis grupos de idades analisados.
- Com base nas medidas obtidas no item (a), tire conclusão sobre o comportamento do peso com o aumento da idade.

Exercício 3 (IME/USP - <http://www.ime.usp/giapaula/cursosgrad.htm>)

Na tabela abaixo estão os dados referentes a uma amostra de 21 trabalhadores em que:

- S: renda (milhares de reais);

- T: tipo de indústria, moderna (M) ou tradicional (T)

- Z: período em que está trabalhando, manhã (M), tarde (T) e noite (N)

| S | T | Z | S | T | Z | S | T | Z |
|-----|---|---|-----|---|---|-----|---|---|
| 4,5 | M | N | 2,7 | T | M | 4,2 | M | M |
| 5 | T | M | 3,5 | T | T | 3,4 | M | N |
| 4,2 | M | M | 3,2 | T | N | 4,4 | M | T |
| 3,7 | M | M | 4,7 | M | N | 3,7 | M | T |
| 3,9 | T | T | 5,5 | M | T | 2,8 | T | M |
| 4,1 | T | N | 4,8 | M | T | 2,5 | T | M |
| 2,9 | T | M | 3,4 | T | M | 2,9 | T | T |

- Agrupe os trabalhadores segundo o tipo de indústria. Calcule para cada grupo a média, mediana variância. Compare os resultados
- Agrupe os trabalhadores segundo o período de trabalho. Calcule para cada grupo a média, mediana variância. Compare os resultados

Exercício 4 (Análise Exploratória de Dados - Prof. Dr. Waldir Medri)

Vinte e uma pacientes de uma clínica médica tiveram seu nível de potássio no plasma medido. Os resultados foram os seguintes:

| Nível | Frequência |
|---------------|------------|
| 2,35 -- 2,55 | 1 |
| 2,55 -- 2,75 | 3 |
| 2,75 -- 2,95 | 2 |
| 2,95 -- 3,15 | 4 |
| 3,15 -- 3,35 | 5 |
| 3,35 -- 3,55 | 6 |

- Determine os quartis: 1º, 2º e 3º
- Construa um histograma

Exercício 5 (Prof. Adriana)

A Tabela a seguir mostra distribuição de frequência para o número de minutos por semana gasto por 400 jovens estudantes do ensino médio assistindo TV. Mais abaixo você encontra dados relativos a uma amostra de 30 estudantes universitários. Verifique se há diferenças notáveis entre os dois grupos estudantis quanto ao número de minutos por semana gasto assistindo TV. Analise as estatísticas descritivas, compare os histogramas e tire sua conclusão.

| Tempo | 300-400 | 400-500 | 500-600 | 600-700 | 700-800 | 800-900 | 900-1000 | 1000-1100 | 1100-1200 |
|------------------|---------|---------|---------|---------|---------|---------|----------|-----------|-----------|
| Nº de Estudantes | 14 | 46 | 58 | 76 | 68 | 62 | 48 | 22 | 6 |

| | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 350 | 560 | 580 | 710 | 945 | 880 | 760 | 640 | 660 | 820 |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

| | | | | | | | | | |
|-----|-----|------|-----|-----|-----|-----|-----|-----|-----|
| 775 | 910 | 920 | 850 | 810 | 790 | 890 | 685 | 730 | 850 |
| 745 | 640 | 1010 | 420 | 770 | 850 | 915 | 840 | 930 | 895 |

Exercício 6 (USP - MAE0219)

Os alunos ingressantes numa universidade matriculados na disciplina de Educação Física foram submetidos a um treinamento de resistência por um período de 2 meses. Antes de iniciarem o treinamento foram submetidos a um teste de resistência quanto ao número de quilômetros que conseguiram correr sem parar. Depois de 2 meses de treinamento, foram novamente submetidos ao mesmo teste. Os dados estão apresentados a seguir.

| Faixas | Frequência | |
|---------|----------------------|-----------------------|
| | Antes do treinamento | Depois do treinamento |
| 0 — 2 | 442 | 80 |
| 2 — 4 | 200 | 200 |
| 4 — 8 | 130 | 280 |
| 8 — 12 | 34 | 179 |
| 12 — 16 | 10 | 43 |
| 16 — 22 | 3 | 8 |

- (a) Construa dois histogramas, um para o teste antes do treinamento e outro para depois. Estime a média, a mediana e os quartis em ambos os casos. Que conclusões você consegue obter?

Exercício 7 (USP - MAE0219)

As rendas mensais em reais dos 25 ingressantes num certo curso de pós-graduação em finanças de uma universidade foram as seguintes (em mil reais): 2,90; 2,90; 2,95; 2,95; 3,10; 3,10; 3,15; 3,20; 3,20; 3,25; 3,30; 3,40; 3,45; 3,45; 3,50; 3,65; 3,65; 3,80; 3,90; 3,90; 4,00; 5,00; 5,20; 5,50 e 6,40. Construa o boxplot. Comente.

Exercício 8 (USP - MAE0219)

Uma indústria, desejando melhorar o nível de seus funcionários em cargos de chefia, montou um curso experimental e indicou 25 funcionários para a primeira turma. Os dados referentes à seção a que pertencem (notas e graus obtidos no curso) estão na tabela a seguir.

- (a) Calcule média, moda, mediana, desvio padrão e coeficiente de variação das variáveis Direito, Política e Estatística.
- (b) Compare e indique as diferenças existentes entre as distribuições das variáveis Direito, Política e Estatística utilizando boxplots.
- (c) Compare o aproveitamento dos funcionários na disciplina Estatística segundo a seção a que eles pertencem.

| 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 |

Exercício 9 (UFRJ)

A tabela a seguir fornece para cada estado do Brasil, o valor total V da produção industrial (em milhões de cruzeiros), o número total (P) de pessoas ocupadas na indústria, o logaritmo decimal de V e o logaritmo decimal de P, segundo o censo industrial de 1980.

| Estado | V | P | $\log_{10}(V)=y$ | $\log_{10}(P)=x$ |
|---------------------|--------|--------|------------------|------------------|
| Amazonas | 333 | 527 | 2,52 | 2,72 |
| Pará | 2655 | 2035 | 3,42 | 3,31 |
| Maranhão | 71 | 271 | 1,85 | 2,43 |
| Piauí | 882 | 1290 | 2,95 | 3,11 |
| Ceará | 8874 | 13776 | 3,95 | 4,14 |
| Rio Grande do Norte | 5989 | 9816 | 3,78 | 3,99 |
| Paraíba | 1469 | 2499 | 3,17 | 3,40 |
| Pernambuco | 9134 | 12720 | 3,96 | 4,10 |
| Alagoas | 924 | 1031 | 2,97 | 3,01 |
| Sergipe | 951 | 961 | 2,98 | 2,98 |
| Bahia | 2234 | 4154 | 3,35 | 3,62 |
| Minas Gerais | 17089 | 30002 | 4,23 | 4,48 |
| Espírito Santo | 2653 | 4402 | 3,42 | 3,64 |
| Rio de Janeiro | 39503 | 49256 | 4,60 | 4,69 |
| São Paulo | 172229 | 195756 | 5,24 | 5,29 |
| Paraná | 4364 | 7619 | 3,64 | 3,88 |
| Santa Catarina | 34335 | 28949 | 4,54 | 4,46 |
| Rio Grande do Sul | 64851 | 91813 | 4,81 | 4,96 |
| Mato Grosso do Sul | 59 | 222 | 1,77 | 2,35 |
| Mato Grosso | 32 | 83 | 1,51 | 1,92 |
| Goiás | 1196 | 2415 | 3,08 | 3,38 |
| Distrito Federal | 99 | 239 | 2,00 | 2,38 |

- Plote o gráfico de dispersão de $\log(V)$ e $\log(P)$ e comente
- Calcule o coeficiente de correlação entre as variáveis de (a)

Exercício 10 (UFRJ)

Em seu livro “*Uncertainties analysis, loads and safety in Structural Engineering* (em português: Análise de incertezas, cargas e segurança em Engenharia de Estruturas), Prentice Hall, 1982”, Gary C. Hart apresenta o conjunto de dados a seguir, que nos permite investigar a relação de dependência entre duas propriedades mecânicas do concreto: X = módulo secante (em 10^6 psi) e Y = força de compressão. (em 10^3 psi)

| | | | | | | | | | | |
|---|------|------|------|------|------|------|------|------|------|------|
| X | 3,41 | 3,52 | 3,57 | 3,61 | 3,43 | 3,59 | 3,62 | 3,56 | 3,35 | 3,47 |
| Y | 8,20 | 7,10 | 7,30 | 8,60 | 6,80 | 7,60 | 8,50 | 6,90 | 5,40 | 6,20 |
| X | 3,53 | 3,33 | 3,54 | 3,22 | 3,49 | 3,25 | 3,79 | 3,64 | 3,67 | 3,72 |
| Y | 7,90 | 5,80 | 9,10 | 4,50 | 6,30 | 5,20 | 9,50 | 8,90 | 7,40 | 8,70 |

- Plote o gráfico de dispersão das variáveis e comente

FAESA Centro Universitário

Prof. Howard Roatti

Análise de Dados Aplicada à Computação

(b) Calcule o coeficiente de correlação entre as variáveis de (a)