

Covariância

O arquivo HorasScore.csv contém os dados relativos às horas de treinamento e a pontuação de um atleta mostrados no exemplo 1.

O código a seguir lê os dados do arquivo na tabela T.

Atribui os valores da coluna Horas para a variável X.

Atribui os valores da coluna Pontos para a variável Y.

Aplica o algoritmo apresentado em sala de aula para calcular R.

Concatena as variáveis X e Y na matriz Z e calcula a matriz de covariância.

```
T = readtable('HorasScore.csv')
```

T = 12x2 table

	Horas	Score
1	9	39
2	15	56
3	25	93
4	14	61
5	10	50
6	18	75
7	0	32
8	16	85
9	5	42
10	19	70
11	16	66
12	20	80

```
n = size(T, 1);           % quantidade de amostras
H = T.Horas;
P = T.Score;
H2 = H.^ 2;              % coluna com valores de H^2
HP = H .* P;             % coluna com valores de H.P
```

```
SHP = (sum(HP) - (sum(H)*sum(P))/n)/n
```

```
SHP = 112.7014
```

```
% Construir a matriz X concatenando H e P
X = [H P];
% Calcula matriz de covariância
cov(X,1)           % o segundo argumento informa para usa a divisão por n (ao invés de n-1)
```

```
ans = 2x2
    43.7431    112.7014
    112.7014    339.2431
```

Coeficiente de Correlação

O arquivo MarketingVendas.csv contém os dados relativos aos gastos com propaganda e vendas de uma empresa mostrados no exemplo 2.

O código a seguir lê os dados do arquivo na tabela T.

Atribui os valores da coluna Marketing para a variável X.

Atribui os valores da coluna Vendas para a variável Y.

Aplica o algoritmo apresentado em sala de aula para calcular R.

Usa função readmatix para ler a matriz Z.

Usa função corr para calcular a calcula a matriz de coeficientes de correlação.

Usa função heatmap para visualizar as correlações.

Coeficiente de correlação de Pearson

```
% Lê os dados para a tabela T
T = readtable('MarketingVendas.csv')
```

```
T = 8x2 table
```

	Marketing	Vendas
1	2.4000	228
2	1.7000	190
3	2	220
4	2.6000	240
5	1.4000	180
6	1.6000	184
7	2.1000	188
8	2.2000	215

```
% Fórmula apresentada em sala de aula
n = size(T, 1); % quantidade de amostras
X = T.Marketing;
Y = T.Vendas;
X2 = X.^2; % coluna com valores de X^2
Y2 = Y.^2; % coluna com valores de Y^2
XY = X.*Y; % coluna com valores X.Y
num = n*sum(XY)-(sum(X)*sum(Y));
```

```
den = sqrt((n*sum(X2)-sum(X)^2))*sqrt((n*sum(Y2)-sum(Y)^2));
R = num/den
```

```
R = 0.8892
```

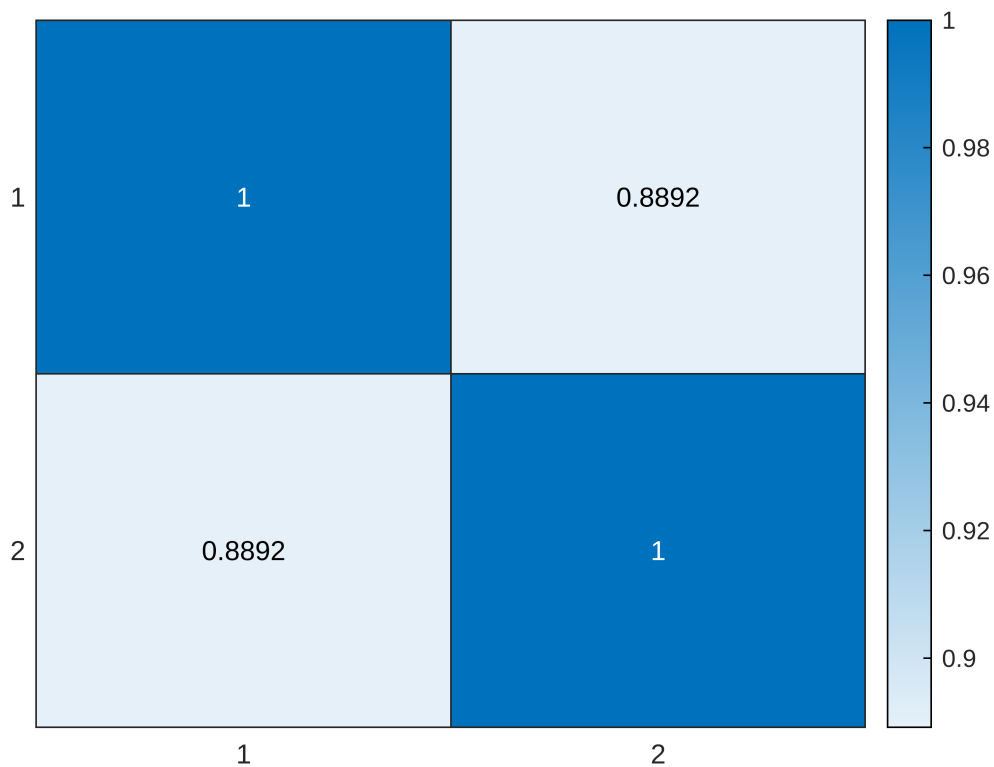
```
% Usa readmatrix para ler diretamente a matriz
Z = readmatrix('MarketingVendas.csv')
```

```
Z = 8x2
    2.4000    228.0000
    1.7000    190.0000
    2.0000    220.0000
    2.6000    240.0000
    1.4000    180.0000
    1.6000    184.0000
    2.1000    188.0000
    2.2000    215.0000
```

```
% Matriz de coeficientes de correlação
R = corr(Z, 'Type', 'Pearson')
```

```
R = 2x2
    1.0000    0.8892
    0.8892    1.0000
```

```
% Mapa de calor
heatmap(R)
```



Coeficiente de correlação de Spearman

```
Ro = corr(Z, 'Type', 'Spearman')
```

```
Ro = 2x2
    1.0000    0.8810
    0.8810    1.0000
```

FORMATIVA

Matriz de covariância e matriz de coeficientes de correlação

O arquivo DadosDataCenter contém 100 medidas de quatro métricas de desempenho de um datacenter, armazenados nas variáveis V1, V2, V3 e V4

Calcular a matriz de covariância, a matriz de coeficientes de correlação e o mapa de calor entre as variáveis.

```
% Lê os dados
matrizFormativa = readmatrix('DadosDataCenter.csv')
```

```
matrizFormativa = 100x4
    38.7937    2.3946    4.7717    0.4250
    17.3294    3.4592    5.2809    0.5688
    15.2657    4.3305   11.5831    0.8884
    33.4756    2.6175    9.7240    0.8419
    24.2862    5.0935    6.9048    0.7473
    11.8213    6.1509   10.2395    0.7537
    22.8005    4.4715   10.3005    0.7918
    23.1420    5.2036    2.4038    0.6296
    19.7583    5.3119    8.8293    0.7828
    21.4044    4.9760    7.4160    0.7370
    ⋮
    ⋮
```

```
tabelaFormativa = readtable('DadosDataCenter.csv')
```

```
tabelaFormativa = 100x4 table
```

	V1	V2	V3	V4
1	38.7937	2.3946	4.7717	0.4250
2	17.3294	3.4592	5.2809	0.5688
3	15.2657	4.3305	11.5831	0.8884
4	33.4756	2.6175	9.7240	0.8419
5	24.2862	5.0935	6.9048	0.7473
6	11.8213	6.1509	10.2395	0.7537
7	22.8005	4.4715	10.3005	0.7918
8	23.1420	5.2036	2.4038	0.6296
9	19.7583	5.3119	8.8293	0.7828
10	21.4044	4.9760	7.4160	0.7370
11	29.2480	5.2441	7.5187	0.7989

	V1	V2	V3	V4
12	32.9526	5.5872	9.8602	0.8982
13	23.6875	4.6508	8.8864	0.7703
14	33.6727	5.4948	7.5117	0.8352
15	20.3425	4.4233	8.2591	0.7209
16	33.4745	4.8507	4.6704	0.4235
17	16.1017	5.3353	4.1417	0.3376
18	17.0421	4.2560	9.7347	0.7284
19	43.4043	1.4437	4.4851	0.3406
20	37.0869	2.8816	8.9525	0.9679
21	23.4083	5.9257	7.9443	0.8168
22	36.6783	3.5935	8.1089	0.7567
23	27.4681	4.8186	8.8118	0.7990
24	11.8938	6.4223	9.7866	0.7580
25	22.3016	3.7435	7.6613	0.6786
26	30.7772	5.2140	6.5561	0.7799
27	20.3016	4.1173	4.9564	0.4172
28	37.2401	0.9857	7.9169	0.4999
29	34.5459	3.6943	10.0600	0.8532
30	22.8401	3.9412	2.2919	0.5530
31	25.7880	4.4349	7.5102	0.7340
32	24.5509	3.0655	8.5631	0.6765
33	27.8997	3.2345	7.2890	0.6713
34	42.3448	2.7893	7.7646	0.8412
35	24.1283	3.8617	5.6240	0.4428
36	14.1158	6.3345	6.9834	0.6972
37	25.2183	4.7583	10.1877	0.8191
38	17.8058	4.4648	5.0413	0.3231
39	48.6086	4.3625	11.0364	0.9283
40	31.7291	5.1780	11.1153	0.9015
41	23.8514	5.0345	11.8510	0.8704
42	19.5186	4.9189	12.8193	0.9629
43	26.8947	4.6192	3.2989	0.2408
44	39.9678	5.5241	12.6500	0.9986

	V1	V2	V3	V4
45	23.7013	4.8500	5.6118	0.4964
46	30.6305	4.9202	9.3120	0.8343
47	29.6937	5.5973	8.9343	0.8582
48	24.3512	4.2860	11.7268	0.8274
49	24.5882	5.3301	8.6704	0.8085
50	33.1111	4.1540	6.7715	0.7368
51	17.6691	4.9173	1.0035	0.1165
52	17.7706	5.2439	9.6502	0.7874
53	19.6764	4.3297	10.8155	0.8781
54	27.5213	4.5496	4.0057	0.4586
55	30.5359	4.7910	9.2899	0.8259
56	8.7786	6.0434	9.5240	0.7592
57	34.5972	4.1997	11.0951	0.8593
58	12.9747	5.7761	7.8411	0.7360
59	16.9373	4.1792	4.6511	0.3908
60	6.7502	7.4707	2.5647	0.4272
61	5.1461	8.9198	6.8122	0.6424
62	16.4514	4.3987	9.9569	0.7383
63	21.1163	4.9511	6.5610	0.5115
64	24.7845	5.1735	8.1872	0.7881
65	18.2528	5.3340	9.0572	0.7802
66	36.0657	4.6917	7.1182	0.6906
67	25.8239	4.7018	6.5546	0.7245
68	37.7837	4.7528	7.8358	0.8207
69	26.0064	4.9799	7.7926	0.7736
70	27.4382	1.7247	3.5353	0.4850
71	26.2422	5.1515	2.6354	0.6503
72	11.0720	6.5047	13.4606	0.8958
73	34.4473	2.2088	5.6807	0.6179
74	25.1192	3.8383	10.1947	0.8663
75	28.8636	5.1058	4.6663	0.7146
76	31.0153	4.0741	8.2064	0.7592
77	23.1861	3.5480	6.3346	0.6380

	V1	V2	V3	V4
78	30.8211	3.1230	7.6426	0.6893
79	25.8356	4.9491	9.6862	0.8203
80	16.5965	5.2379	10.1702	0.7926
81	36.1462	5.6765	7.4520	0.5558
82	26.7100	4.1592	2.8358	0.6015
83	37.8570	5.2057	6.5117	0.8123
84	22.6608	3.8657	8.3652	0.7060
85	20.8642	4.1809	9.2126	0.7352
86	30.0391	4.9479	8.3923	0.8089
87	3.8179	5.3333	6.2348	0.5763
88	0	4.6220	6.8908	0.3908
89	47.4267	3.3906	5.4668	0.4369
90	14.5630	3.8605	5.2501	0.5714
91	22.5422	4.3990	8.8830	0.7492
92	10.4782	6.0648	9.7637	0.7821
93	43.1718	5.2199	9.3028	0.9094
94	33.6020	3.1594	7.4684	0.7007
95	25.8525	4.6341	7.8827	0.7554
96	19.5939	2.9531	9.4092	0.7625
97	27.5616	4.1807	9.1231	0.7712
98	15.3278	3.8151	4.2802	0.5491
99	28.9378	4.5276	15.0390	0.9725
100	36.8700	5.6936	11.8123	0.9438

```
% Calcula matriz de covariância
```

```
n = size(tabelaFormativa, 1); % quantidade de amostras
```

```
H = tabelaFormativa.V1;
```

```
P = tabelaFormativa.V2;
```

```
J = tabelaFormativa.V3;
```

```
L = tabelaFormativa.V4;
```

```
A = [H P J L]
```

```
A = 100x4
```

```

38.7937    2.3946    4.7717    0.4250
17.3294    3.4592    5.2809    0.5688
15.2657    4.3305   11.5831    0.8884
33.4756    2.6175    9.7240    0.8419
24.2862    5.0935    6.9048    0.7473

```

```

11.8213    6.1509    10.2395    0.7537
22.8005    4.4715    10.3005    0.7918
23.1420    5.2036    2.4038    0.6296
19.7583    5.3119    8.8293    0.7828
21.4044    4.9760    7.4160    0.7370
⋮

```

```
cov(A, 1)
```

```

ans = 4x4
88.6678    -4.9893    1.5050    0.3484
-4.9893    1.3593    0.4590    0.0417
1.5050     0.4590    7.0699    0.3816
0.3484     0.0417    0.3816    0.0306

```

```
% Calcula matriz de coeficientes de correlação de Pearson
```

```
R = corr(matrizFormativa, 'Type', 'Pearson')
```

```

R = 4x4
1.0000    -0.4545    0.0601    0.2114
-0.4545    1.0000    0.1481    0.2042
0.0601     0.1481    1.0000    0.8199
0.2114     0.2042    0.8199    1.0000

```

```
Ro = corr(matrizFormativa, 'Type', 'Spearman')
```

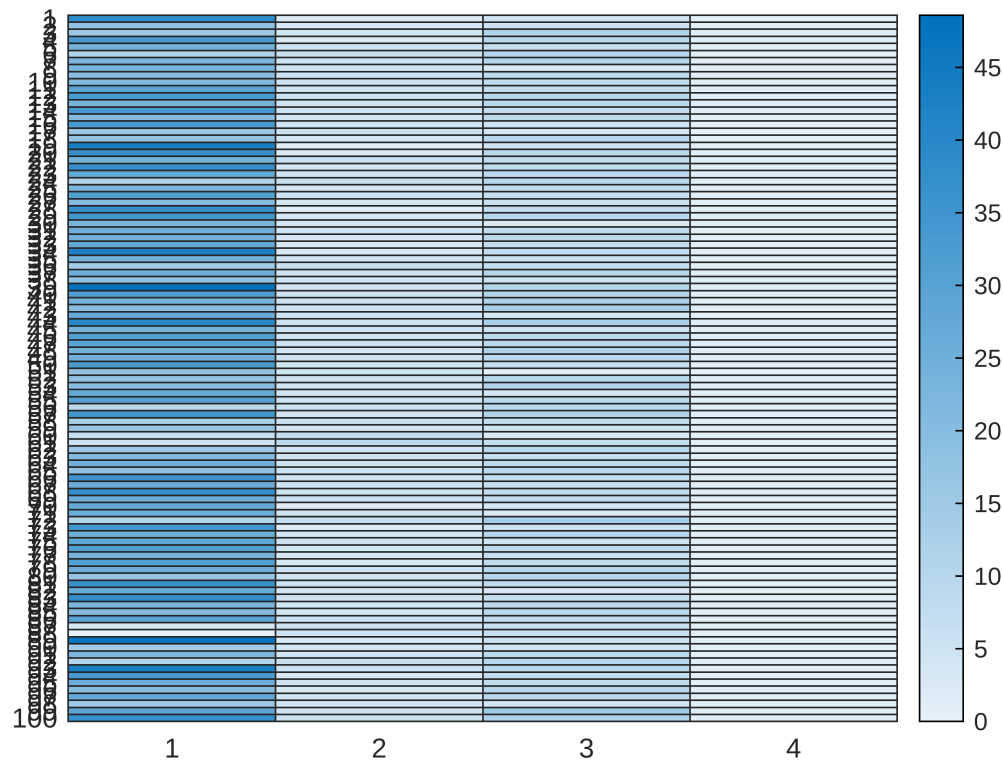
```

Ro = 4x4
1.0000    -0.3113    0.0165    0.2621
-0.3113    1.0000    0.1759    0.2577
0.0165     0.1759    1.0000    0.8403
0.2621     0.2577    0.8403    1.0000

```

```
% Mapa de calor da matriz de correlação
```

```
heatmap(matrizFormativa)
```

Com base nos resultados dos coeficientes de correlação de Spearman responder:

Quais variáveis apresentam correlação negativa entre si? (responder aqui)

Possuem a mesma

Quais as duas variáveis que possuem maior correlação positiva? (responder aqui)

Possuem a mesma

Quais as duas variáveis que são menos correlacionadas? (responder aqui)

Sperman

Formativa

O escritório de projetos deseja identificar se existe corelação entre a estimativa de horas e o esforço em horas despendido em seu desenvolvimento. Os dados utilizados encontram-se no arquivo “AMP_esforco_sprint.csv”.

Avaliação de correlação

```
T = readtable("AMP_esforco_sprint.csv")
```

T = 80x5 table

...

	ID_SPURT	Estimativa_UCP	Estimativa_Horas	Realizado_Horas
1	1	16.7437	91.8900	84.7000
2	2	41.3325	225.7400	194
3	3	26.0250	111.6500	94
4	4	11.4835	49.6900	41
5	5	44.3278	261.5300	217
6	6	49.2950	283.8600	219.9000
7	7	23.5032	101.0300	85.6000
8	8	43.7396	186.9900	259.9000
9	9	39.8510	171.3100	230.1000
10	10	17.1645	130.8000	119.3000
11	11	15.9470	68.5500	68
12	12	66.5514	284.0800	262
13	13	22.1351	95.1500	85
14	14	10.3950	44.6900	35
15	15	61.5668	264.6600	245.9000
16	16	3.7973	16.3200	22.1700
17	17	23.3811	100.5100	116
18	18	2.4077	10.3500	10.8300
19	19	25.6640	110.3200	122.8000
20	20	39.6039	170.2500	166.3200
21	21	8.3113	135.7300	140.1700
22	22	25.1214	107.9900	129.6500
23	23	57.0634	288.2900	217.5000
24	24	9.8040	42.1500	45.6000
25	25	20.2406	87.0100	62.6000
26	26	26.6305	114.4800	129.8000
27	27	15.0453	84.6800	64.4000
28	28	9.4157	40.4800	52.3300
29	29	42.2704	173.6600	234.4000
30	30	15.3824	66.1300	64.7000
31	31	43.9759	189.0400	150
32	32	4.7349	20.3500	20
33	33	35.4713	152.4800	142.5000

	ID_SPURT	Estimativa_UCP	Estimativa_Horas	Realizado_Horas
34	34	37.3295	160.4700	211.4300
35	35	6.8153	29.3000	35
36	36	12.0159	51.6500	31.6000
37	37	21.0622	90.5400	86.3000
38	38	8.5145	43.6000	24.7300
39	39	47.0685	217.2800	256.1200
40	40	26.9838	116	143.7000
41	41	26.8078	115.2400	151.4200
42	42	5.6851	94.4400	72.9200
43	43	36.3884	156.4300	192
44	44	4.8065	20.6600	18.5000
45	45	6.3179	117.1600	92.7000
46	46	14.7574	63.4400	76.7000
47	47	50.3940	173.6400	200.9300
48	48	6.9571	49.9100	35
49	49	45.2938	214.7100	163.6700
50	50	44.6338	191.8700	241.3800
51	51	64.6212	277.7900	238.3700
52	52	53.9129	226.7000	218
53	53	36.3043	156.0600	148.4700
54	54	17.4482	75.0100	59.6700
55	55	16.2258	69.7500	73.3300
56	56	26.0273	111.8900	109
57	57	8.1363	34.9800	32.6500
58	58	30.1959	129.8100	129
59	59	9.0114	38.7400	42.4800
60	60	5.5682	63.9400	42.1700
61	61	12.8395	55.1900	83
62	62	45.9896	154.7100	173.5800
63	63	11.2611	48.4100	47.5000
64	64	4.4002	18.9200	16.1000
65	65	38.1989	164.2100	115.5800
66	66	56.4599	242.7100	244.8500

	ID_SPURT	Estimativa_UCP	Estimativa_Horas	Realizado_Horas
67	67	5.4351	53.3600	35.5800
68	68	51.3956	220.2500	219.6500
69	69	13.6361	58.6200	49.7500
70	70	41.1542	176.9100	153
71	71	48.4944	208.4700	182.6000
72	72	7.2325	31.0900	28.6000
73	73	5.9796	65.7000	50.7800
74	74	7.0998	30.5200	33.6700
75	75	15.1255	65.0200	40.7300
76	76	26.3983	113.4800	83.4200
77	77	11.8821	59.9200	68
78	78	22.9869	98.8200	84.5800
79	79	30.4373	130.8400	138.4000
80	80	11.2890	48.5300	58.6000

```
matrizT = readmatrix("AMP_esforco_sprint.csv")
```

```
matrizT = 80x5
    1.0000    16.7437    91.8900    84.7000    1.0000
    2.0000    41.3325   225.7400   194.0000         0
    3.0000    26.0250   111.6500    94.0000    1.0000
    4.0000    11.4835    49.6900    41.0000    1.0000
    5.0000    44.3278   261.5300   217.0000         0
    6.0000    49.2950   283.8600   219.9000    1.0000
    7.0000    23.5032   101.0300    85.6000    1.0000
    8.0000    43.7396   186.9900   259.9000    1.0000
    9.0000    39.8510   171.3100   230.1000    1.0000
   10.0000    17.1645   130.8000   119.3000    1.0000
      ⋮
```

Plotar o diagrama de dispersão e calcular o coeficiente de correlação.

Dicas:

- Salvar as estimativas de horas no vetor X
- Salvar quantidade de horas realizadas em horas no vetor Y
- Usar o comando scatter do Matlab para plotar o diagrama de dispersão
- Concatenar os vetores X e Y na matriz Z
- Calcular R usando a função corrcoef do MatLab passando Z como argumento
- Calcular o mapa de calor da matriz de coeficientes de correlação

```
quantidadeAmostra = size(T,1)
```

```
quantidadeAmostra = 80
```

```

% Salvar as estimativas de horas no vetor X
X = T.Estimativa_Horas;
% Salvar o realizado em horas no vetor Y
Y = T.Realizado_Horas;

% Calcular a matriz de coeficientes de correlação de Pearson
R = corr(matrizT, 'Type', 'Pearson')

```

```

R = 5x5
    1.0000    -0.1202    -0.1886    -0.1973    -0.1422
   -0.1202     1.0000     0.9493     0.9324    -0.0846
   -0.1886     0.9493     1.0000     0.9381    -0.0960
   -0.1973     0.9324     0.9381     1.0000    -0.0813
   -0.1422    -0.0846    -0.0960    -0.0813     1.0000

```

```

% Calcular o mapa de calor da matriz de coeficientes de correlação
heatmap(matrizT)

```

