

## 2ª Lista de Exercícios - Análise Bidimensional

**Exercício 1.** Numa pesquisa sobre rotatividade de mão-de-obra, para uma amostra de 40 pessoas foram observadas duas variáveis: número de empregos nos últimos dois anos ( $X$ ) e salário mais recente, em número de salários mínimos ( $Y$ ). Os resultados foram:

Indivíduo	$X$	$Y$	Indivíduo	$X$	$Y$
1	1	6	21	2	4
2	3	2	22	3	2
3	2	4	23	4	1
4	3	1	24	1	5
5	2	4	25	2	4
6	2	1	26	3	2
7	3	3	27	4	1
8	1	5	28	1	5
9	2	2	29	4	4
10	3	2	30	3	3
11	2	5	31	2	2
12	3	2	32	1	1
13	1	6	33	4	1
14	2	6	34	2	6
15	3	2	35	4	2
16	4	2	36	3	1
17	1	5	37	1	4
18	2	5	38	3	2
19	2	1	39	2	3
20	2	1	40	2	5

- Usando a mediana, classifique os indivíduos em dois níveis, alto e baixo, para cada uma das variáveis, e construa a distribuição de frequências conjunta das duas classificações.
- Qual a porcentagem das pessoas com baixa rotatividade e ganhando pouco?
- Qual a porcentagem das pessoas que ganham pouco?
- Entre as pessoas com baixa rotatividade, qual a porcentagem das que ganham pouco?
- A informação adicional dada em (d) mudou muito a porcentagem observada em (c)? O que isso significa?
- Utilizando frequências relativas e/ou um gráfico de barras, verifique se há relações entre as variáveis rotatividade e salário.
- Encontre o valor de  $\chi^2$ , do coeficiente de contingência  $C$  e do coeficiente  $T$ . Interprete.

**Exercício 2.** Uma companhia de seguros analisou a frequência com que 2.000 segurados (1.000 homens e 1.000 mulheres) usaram o hospital. Os resultados foram:

	Homens	Mulheres
Usaram o hospital	100	150
Não usaram o hospital	900	850

- Calcule a proporção de homens entre os indivíduos que usaram o hospital.
- Calcule a proporção de homens entre os indivíduos que não usaram o hospital.
- O uso do hospital independe do sexo do segurado?
- Se achar necessário, utilize um gráfico de barras para visualizar se há ou não relação entre o gênero e o uso do hospital.
- Por fim, encontre o valor de  $\chi^2$ , do coeficiente de contingência  $C$  e do coeficiente  $T$ . Interprete.

**Exercício 3.** A Companhia A de dedetização afirma que o processo por ela utilizado garante um efeito mais prolongado do que aquele obtido por seus concorrentes mais diretos. Uma amostra de vários ambientes dedetizados foi colhida e anotou-se a duração do efeito de dedetização. Os resultados estão na tabela abaixo. Você acha que existe alguma evidência a favor ou contra a afirmação feita pela Companhia A?

Companhia	Duração do efeito de dedetização		
	Menos de 4 meses	De 4 a 8 meses	Mais de 8 meses
A	64	120	16
B	104	175	21
C	27	48	5

**Exercício 4.** Para cada par de variáveis abaixo, diga se você espera uma dependência linear e nos casos afirmativos avalie o coeficiente de correlação. Como ficaria o diagrama de dispersão em cada caso?

- Peso e altura dos alunos do primeiro ano de um curso de Administração.
- Peso e altura dos funcionários de um escritório.
- Quantidade de trigo produzida e quantidade de água recebida por canteiros numa estação experimental.
- Notas de Cálculo e Estatística de uma classe onde as duas disciplinas são lecionadas.
- Acuidade visual e idade de um grupo de pessoas.
- Renda familiar e porcentagem dela gasta em alimentação.
- Número de peças montadas e resultado de um teste de inglês por operário.

**Exercício 5.** A Tabela abaixo mostra o número de anos de serviço ( $X$ ) e o número de clientes de 10 agentes de uma companhia de seguros. Estamos interessados em verificar se existe algum tipo de relação entre essas duas variáveis numéricas.

Agente	$X$	$Y$
A	2	48
B	3	50
C	4	56
D	5	52
E	4	43
F	6	60
G	7	62
H	8	58
I	8	64
J	10	72

- Construa um diagrama de dispersão para essas variáveis. Há indícios de relação entre elas? De qual tipo e grau?
- Calcule o coeficiente de correlação linear e verifique se os indícios levantados acima estão de acordo com o valor obtido.

**Exercício 6.** Pensando em estabelecer públicos alvo para determinadas políticas públicas, uma Secretaria de Saúde deseja saber se existe relação entre a renda de uma família e o quanto se gasta com saúde. Para isso, foram coletadas informações de 10 famílias, observando-se a renda bruta mensal ( $X$ ) e a porcentagem da renda gasta em saúde ( $Y$ ). Os dados estão dispostos na tabela abaixo:

Família	$X$	$Y$
A	12	7,2
B	16	7,4
C	18	7,0
D	20	6,5
E	28	6,6
F	30	6,7
G	40	6,0
H	48	5,6
I	50	6,0
J	54	5,5

- Construa um diagrama de dispersão para essas variáveis. Há indícios de relação entre elas? De qual tipo e grau?
- Calcule o coeficiente de correlação linear e verifique se os indícios levantados acima estão de acordo com o valor obtido.

**Exercício 7.** Oito indivíduos foram submetidos a um teste sobre conhecimento de língua estrangeira e, em seguida, mediu-se o tempo gasto para cada um aprender a operar uma determinada máquina. As variáveis medidas foram:

$X$ : resultado obtido no teste (máximo = 100 pontos);

$Y$ : tempo, em minutos, necessário para operar a máquina satisfatoriamente.

Os dados obtidos estão apresentados na tabela abaixo

Indivíduo	$X$	$Y$
A	45	343
B	52	368
C	61	355
D	70	334
E	74	337
F	76	381
G	80	345
H	90	375

- Construa um diagrama de dispersão para essas variáveis. Há indícios de relação entre elas? De qual tipo e grau?
- Calcule o coeficiente de correlação linear e verifique se os indícios levantados acima estão de acordo com o valor obtido.

**Exercício 8.** Abaixo estão os dados referentes à porcentagem da população economicamente ativa empregada no setor primário e o respectivo índice de analfabetismo para algumas regiões metropolitanas brasileiras.

Regiões metropolitanas	Setor primário	Índice de analfabetismo
São Paulo	2,0	17,5
Rio de Janeiro	2,5	18,5
Belém	2,9	19,5
Belo Horizonte	3,3	22,2
Salvador	4,1	26,5
Porto Alegre	4,3	16,6
Recife	7,0	36,6
Fortaleza	13,00	38,4

- Faça o diagrama de dispersão.
- Você acha que existe uma dependência linear entre as duas variáveis?
- Calcule o coeficiente de correlação.
- Existe alguma região com comportamento diferente das demais? Se existe, elimine o valor correspondente e recalcule o coeficiente de correlação.

**Exercício 9.** Quer se verificar a relação entre o tempo de reação e o número de alternativas apresentadas a indivíduos acostumados a tomadas de decisão. Planejou-se um experimento em que se pedia ao participante para classificar objetos segundo um critério previamente discutido. Participaram do experimento 15 executivos divididos aleatoriamente em grupos de cinco. Pediu-se, então, a cada grupo para classificar dois, três e quatro objetos, respectivamente. Os dados estão abaixo.

Nº de objetos	2	3	4
Tempo de reação	1, 2, 3, 3, 4	2, 3, 4, 4, 5	4, 5, 5, 6, 7

- (a) Faça o gráfico de dispersão das duas variáveis.
- (b) Qual o coeficiente de correlação entre elas?

**Exercício 10.** A tabela a seguir mostra dados da superfície (em  $km^2$ ), população estimada e densidade ( $hab/km^2$ ) das unidades federativas (UF) do Brasil, por região, em 2007 (IBGE).

Regiões	UF	Superfície	População	Densidade
Norte	RO	237.576,167	1.453.756	6,12
	AC	164.165,250	653.385	3,99
	AM	1.559.161,810	3.221.940	2,07
	RR	224.298,980	395.725	1,76
	PA	1.247.689,515	7.065.573	5,66
	AP	142.814,585	587.311	4,11
	TO	277.620,914	1.243.627	4,48
	Subtot.	3.853.327	14.623.317	3,79
Nordeste	MA	331.983,293	6.118.995	18,43
	PI	251.529,186	3.032.435	12,06
	CE	148.825,602	8.185.250	55,0
	RN	52.796,791	3.013.740	57,08
	PB	56.439,838	3.641.397	64,52
	PE	98.311,616	8.486.638	86,32
	AL	27.767,661	3.037.231	108,38
	SE	21.910,348	1.939.426	88,52
	BA	564.692,669	14.080.670	24,94
	Subtot.	1.554.257	51.535.782	33,16
Sudeste	MG	586.528,293	19.273.533	32,86
	ES	46.077,519	3.351.669	72,74
	RJ	43.696,054	15.420.450	352,90
	SP	248.209,426	39.827.690	160,46
	Subtot.	924.511	77.873.342	84,23
Sul	PR	199.314,850	10.284.503	51,6
	SC	95.346,181	5.866.487	61,53
	RS	281.748,583	10.582.287	37,56
	Subtot.	576.410	26.733.877	46,38
Centro-Oeste	MS	357.124,962	2.265.813	6,34
	MT	903.357,908	2.854.642	3,16
	GO	340.086,698	5.647.035	16,6
	DF	5.801,937	2.455.903	423,29
	Subtot.	1.606.372	13.223.393	8,23
Brasil	Total	8.514.876,599	183.989.711	21,61

Análise a população total do Brasil (em 2007), segundo as regiões geográficas. Existe associação entre essas variáveis? Utilize gráficos e algum índice de correlação para suas conclusões.

**Exercício 11.** Faça o gráfico  $q \times q$  para os dois conjuntos de dados em  $A$  e  $B$  a seguir.

A	65	54	49	60	70	25	87	100	70	102	40	47
B	48	35	45	50	52	20	72	102	46	82	-	-

- Verifique, a través do gráfico  $q \times q$  se o comportamento dessas duas variáveis é similar em algum sentido.
- Compare os boxplots dessas duas variáveis para auxiliar na análise.

**Exercício 12.** O departamento de vendas de certa companhia foi formado há um ano com a admissão de 15 vendedores. Nessa época, foram observados para cada um dos vendedores os valores de três variáveis:

$T$ : resultado em um teste apropriado para vendedores;

$E$ : anos de experiência de vendas;

$G$ : conceito do gerente de venda, quanto ao currículo do candidato.

O diretor da companhia resolveu agora ampliar o quadro de vendedores e pede sua colaboração para responder a algumas perguntas. Para isso, ele lhe dá informações adicionais sobre duas variáveis:

$V$ : volume médio mensal de vendas em salários mínimos;

$Z$ : zona da capital para a qual o vendedor foi designado.

O quadro de resultados é o seguinte:

Vendedor	$T$ : teste	$E$ : experiência	$G$ : conceito do gerente	$V$ : vendas	$Z$ : zona
1	8	5	Bom	54	Norte
2	9	2	Bom	50	Sul
3	7	2	Mau	48	Sul
4	8	1	Mau	32	Oeste
5	6	4	Bom	30	Sul
6	8	4	Bom	30	Oeste
7	5	3	Bom	29	Norte
8	5	3	Bom	27	Norte
9	6	1	Mau	24	Oeste
10	7	3	Mau	24	Oeste
11	4	4	Bom	24	Sul
12	7	2	Mau	23	Norte
13	3	3	Mau	21	Sul
14	5	1	Mau	21	Oeste
15	3	2	Bom	16	Norte

Mais especificamente, o diretor lhe pede que responda aos sete itens seguintes:

- Faça o histograma da variável  $V$  em classes de 10, tendo por limite inferior da primeira classe o valor 15.
- Encontre a média e a variância da variável  $V$ . Suponha que um vendedor seja considerado excepcional se seu volume de vendas é dois desvios padrões superior à média geral. Quantos vendedores excepcionais existem na amostra?
- O diretor de vendas anunciou que transferirá para outra praça todos os vendedores cujo volume de vendas for inferior ao 1º quartil da distribuição. Qual o volume mínimo de vendas que um vendedor deve realizar para não ser transferido?

- (d) Os vendedores argumentam com o diretor que esse critério não é justo, pois há zonas de venda privilegiadas. A quem você daria razão?
- (e) Qual das três variáveis observadas na admissão do pessoal é mais importante para julgar um futuro candidato ao emprego?
- (f) Qual o grau de associabilidade entre o conceito do gerente e a zona a que o vendedor foi designado? Você tem explicação para esse resultado?
- (g) Qual o grau de associação entre o conceito do gerente e o resultado do teste? E entre zona e vendas?

**Exercício 13. (Opcional/Interessante)** A seção de assistência técnica da Companhia MB tem cinco funcionários: A, B, C, D e E, cujos tempos de serviço na companhia são, respectivamente, um, três, cinco, cinco e sete anos.

- (a) Faça um gráfico representando a distribuição de frequência dos tempos de serviço  $X$ .
- (b) Calcule a média  $me(X)$ , a variância  $var(X)$  e a mediana  $md(X)$ .

Duas novas firmas, a Verde e a Azul, solicitaram o serviço de assistência técnica da MB. Um mesmo funcionário pode ser designado para atender a ambos os pedidos, ou dois funcionários podem fazê-lo. Assim, o par  $(A, B)$  significa que o funcionário A atenderá à firma Verde e o funcionário B, à firma Azul.

- (c) Escreva os 25 possíveis pares de funcionários para atender a ambos os pedidos.
- (d) Para cada par, calcule o tempo médio de serviço  $\bar{X}$ , faça a distribuição de frequência e uma representação gráfica. Compare com o resultado de (a).
- (e) Calcule para os 25 valores de  $\bar{X}$  os parâmetros  $me(\bar{X})$ ,  $var(\bar{X})$  e  $md(\bar{X})$ . Compare com os resultados obtidos em (b). Que tipo de conclusão você poderia tirar?
- (f) Para cada par obtido em (c), calcule a variância do par e indique-a por  $S^2$ . Faça a representação gráfica da distribuição dos valores de  $S^2$ .
- (g) Calcule  $me(S^2)$  e  $var(S^2)$ .
- (h) Indicando por  $X_1$  a variável que expressa o tempo de serviço do funcionário que irá atender à firma Verde e  $X_2$  o que irá atender à firma Azul, faça a distribuição conjunta da variável bidimensional  $(X_1, X_2)$ .
- (i) As duas variáveis  $X_1$  e  $X_2$  são independentes?
- (j) O que você pode falar sobre as distribuições marginais de  $X_1$  e  $X_2$ ?
- (l) Suponha agora que três firmas solicitem o serviço de assistência técnica. Quantas triplas podem ser formadas?
- (m) Sem calcular todas as possibilidades, como você acha que ficaria o histograma de  $\bar{X}$ ? E  $me(\bar{X})$ ? E  $var(\bar{X})$ ?
- (n) E sobre a variável  $S^2$ ?
- (o) A variável tridimensional  $(X_1, X_2, X_3)$  teria alguma propriedade especial para as suas distribuições marginais?