

Universidade Federal do Rio Grande do Norte

Centro de Ciências Exatas e da Terra – Departamento de Estatística

Disciplina: Estatística Aplicada a engenharia I (EST0323)

Professora: Anna Rafaella da S. Marinho

Aluno(a):

Lista I

1. Considere a tabela de dados brutos a seguir, referente a informações de um questionário aplicado a alunos do primeiro ano de uma escola contendo as seguintes informações:

- Id: Identificação do aluno;
- Turma: turma em que o aluno foi alocado (A ou B);
- Sexo: F se feminino, M se masculino;
- Idade: idade em anos;
- Alt: altura em metros;
- Peso: peso em quilogramas;
- Filhos: número de filhos na família;
- Fuma: Hábito de fumar (Sim ou Não);
- Toler: tolerância ao cigarro (I– indiferente, P – incomoda pouco, M – incomoda muito);
- Exerc: horas de atividade física por semana;
- Cine: número de vezes em que vai ao cinema por semana;
- OpCine: opinião a respeito das salas de cinema da cidade (B – regular ou boa, M – muito boa);
- TV: horas gastas assistindo televisão, por semana;
- OpTV: opinião a respeito da qualidade da programação de TV (R – ruim; M – média, B – boa, N – não sabe).

- a) A partir dos dados da tabela, construa histogramas para as variáveis contínuas, escolha e faça uma representação gráfica adequada para cada variável discreta ou qualitativa.
- b) Construa uma tabela de frequências para a variável peso e obtenha os percentis de 35% e de 70%. Interprete-os.

Tabela 1.1: Informações de questionário estudantil - dados brutos.

Id	Turma	Sexo	Idade	Alt	Peso	Filh	Fuma	Toler	Exer	Cine	OpCine	TV	OpTV
1	A	F	17	1,60	60,5	2	NAO	P	0	1	B	16	R
2	A	F	18	1,69	55,0	1	NAO	M	0	1	B	7	R
3	A	M	18	1,85	72,8	2	NAO	P	5	2	M	15	R
4	A	M	25	1,85	80,9	2	NAO	P	5	2	B	20	R
5	A	F	19	1,58	55,0	1	NAO	M	2	2	B	5	R
6	A	M	19	1,76	60,0	3	NAO	M	2	1	B	2	R
7	A	F	20	1,60	58,0	1	NAO	P	3	1	B	7	R
8	A	F	18	1,64	47,0	1	SIM	I	2	2	M	10	R
9	A	F	18	1,62	57,8	3	NAO	M	3	3	M	12	R
10	A	F	17	1,64	58,0	2	NAO	M	2	2	M	10	R
11	A	F	18	1,72	70,0	1	SIM	I	10	2	B	8	N
12	A	F	18	1,66	54,0	3	NAO	M	0	2	B	0	R
13	A	F	21	1,70	58,0	2	NAO	M	6	1	M	30	R
14	A	M	19	1,78	68,5	1	SIM	I	5	1	M	2	N
15	A	F	18	1,65	63,5	1	NAO	I	4	1	B	10	R
16	A	F	19	1,63	47,4	3	NAO	P	0	1	B	18	R
17	A	F	17	1,82	66,0	1	NAO	P	3	1	B	10	N
18	A	M	18	1,80	85,2	2	NAO	P	3	4	B	10	R
19	A	F	20	1,60	54,5	1	NAO	P	3	2	B	5	R
20	A	F	18	1,68	52,5	3	NAO	M	7	2	B	14	M
21	A	F	21	1,70	60,0	2	NAO	P	8	2	B	5	R
22	A	F	18	1,65	58,5	1	NAO	M	0	3	B	5	R
23	A	F	18	1,57	49,2	1	SIM	I	5	4	B	10	R
24	A	F	20	1,55	48,0	1	SIM	I	0	1	M	28	R
25	A	F	20	1,69	51,6	2	NAO	P	8	5	M	4	N
26	A	F	19	1,54	57,0	2	NAO	I	6	2	B	5	R
27	B	F	23	1,62	63,0	2	NAO	M	8	2	M	5	R
28	B	F	18	1,62	52,0	1	NAO	P	1	1	M	10	R
29	B	F	18	1,57	49,0	2	NAO	P	3	1	B	12	R
30	B	F	25	1,65	59,0	4	NAO	M	1	2	M	2	R
31	B	F	18	1,61	52,0	1	NAO	P	2	2	M	6	N
32	B	M	17	1,71	73,0	1	NAO	P	1	1	B	20	R
33	B	F	17	1,65	56,0	3	NAO	M	2	1	B	14	R
34	B	F	17	1,67	58,0	1	NAO	M	4	2	B	10	R
35	B	M	18	1,73	87,0	1	NAO	M	7	1	B	25	B
36	B	F	18	1,60	47,0	1	NAO	P	5	1	M	14	R
37	B	M	17	1,70	95,0	1	NAO	P	10	2	M	12	N
38	B	M	21	1,85	84,0	1	SIM	I	6	4	B	10	R
39	B	F	18	1,70	60,0	1	NAO	P	5	2	B	12	R
40	B	M	18	1,73	73,0	1	NAO	M	4	1	B	2	R
41	B	F	17	1,70	55,0	1	NAO	I	5	4	B	10	B
42	B	F	23	1,45	44,0	2	NAO	M	2	2	B	25	R
43	B	M	24	1,76	75,0	2	NAO	I	7	0	M	14	N
44	B	F	18	1,68	55,0	1	NAO	P	5	1	B	8	R
45	B	F	18	1,55	49,0	1	NAO	M	0	1	M	10	R
46	B	F	19	1,70	50,0	7	NAO	M	0	1	B	8	R
47	B	F	19	1,55	54,5	2	NAO	M	4	3	B	3	R
48	B	F	18	1,60	50,0	1	NAO	P	2	1	B	5	R
49	B	M	17	1,80	71,0	1	NAO	P	7	0	M	14	R
50	B	M	18	1,83	86,0	1	NAO	P	7	0	M	20	B

- c) Divida a variável idade em três categorias (menores de 18 anos, idade entre 18 e 21 inclusive, e maiores de 21 anos). Construa um gráfico de barras para essa variável, incluindo todos os indivíduos e um para cada sexo. Interprete os resultados obtidos.
- d) Utilize medidas de resumo para sumarizar todas as variáveis quantitativas.
2. (Bussab & Morettin - 2002) O departamento pessoal de uma certa firma fez um levantamento dos salários dos 120 funcionários do setor administrativo, obtendo os resultados (em salários mínimos) da tabela abaixo.

Faixa Salarial	Frequência Relativa
[0, 2)	0.25
[2, 4)	0.40
[4, 6)	0.20
[6, 10)	0.15

- a) Esboce o histograma correspondente;
- b) Calcule a média, a variância, o desvio padrão e o coeficiente de variação.
- c) Calcule o primeiro quartil e a mediana.
- d) Construa um box plot. O que você pode concluir a partir dele?
- e) Se for concedido um aumento de 100% para todos os 120 funcionários, haverá alteração na média? E na variância? Justifique sua resposta.
- f) Se for concedido um abono de dois salários mínimos para todos os funcionários, haverá alteração na média? E na variância? E na mediana? E no coeficiente de variação? Justifique sua resposta.
3. 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 números de salários mínimos (Y). Os resultados são apresentados na tabela a seguir.
- a) Usando a mediana, classifique os indivíduos em dois níveis, alto e baixo, para cada uma das variáveis, e construa a distribuição de frequências para as duas variáveis.
- b) Qual o percentual de pessoas tem baixa rotatividade e ganham pouco?
- c) Qual é o percentual de pessoas que ganham pouco?
- d) Entre as pessoas com baixa rotatividade, qual é o percentual que ganha pouco?

- e) A informação adicional dada em (d) mudou muito o percentual observado em (c)? O que isso significa?
- f) Construa um gráfico de dispersão e analise a associação entre as variáveis.
- g) Calcule o coeficiente de correlação. Baseado nesse número você diria que existe dependência entre as duas variáveis? Justifique.

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

4. Um professor está interessado em saber se existe alguma relação entre o número de faltas e a nota de cada aluno. Com esse objetivo, colheu uma amostra e anotou os dados na tabela a seguir:

Nº de Faltas	8	2	5	12	15	9	6
Nota Final	78	92	90	58	43	74	81

Dados:

$$\sum_{i=1}^7 X_i = 57, \sum_{i=1}^7 Y_i = 516, \sum_{i=1}^7 X_i Y_i = 3751, \sum_{i=1}^7 X_i^2 = 579, \sum_{i=1}^7 Y_i^2 = 39898$$

Em que:

X:= número de faltas;

Y:= nota final;

- a) Construa o diagrama de dispersão. Com base no diagrama, as variáveis **Número de Faltas** e **Nota Final** estão relacionadas? Em caso afirmativo, qual é o tipo de relação? Justifique.
- b) Calcule o coeficiente de correlação e prove se a sua resposta ao item (a) está correta. Qual é o tipo de correlação entre as variáveis?
5. Em uma cidade onde se publicam três jornais A, B e C, constatou-se que entre 1000 famílias, 470 assinam A, 420 assinam B, 315 assinam C, 110 A e B, 220 A e C, 140 B e C, e, 75 assinam os três. escolhendo-se uma família ao acaso, qual a probabilidade de que ela:
- a) não assine nenhum dos três jornais?
- b) assine apenas um dos três jornais?
- c) assine pelo menos dois jornais?
6. Num período de um mês, 100 pacientes sofrendo de determinada doença foram internados em um hospital. Informações sobre o método de tratamento aplicado em cada paciente e o resultado final obtido estão na tabela abaixo:

Resultado\Tratamento	A	B	Soma
Cura total	24	16	40
Cura parcial	24	16	40
Morte	12	8	20
Soma	60	40	100

- a) Sorteando aleatoriamente um desses pacientes, determine a probabilidade de o paciente escolhido:
- ter sido submetido ao tratamento A;
 - ter sido totalmente curado;
 - ter sido submetido ao tratamento A e ter sido parcialmente curado;
 - ter sido submetido ao tratamento A ou ter sido parcialmente curado;
- b) Sorteando dois dos pacientes, qual a probabilidade de que:
- tenham recebido tratamentos diferentes?
 - pelo menos um deles tenha sido curado totalmente?

7. Num certo colégio, 4% dos homens e 1% das mulheres têm mais de $1,75m$ de altura. 60% dos estudantes são mulheres. Um estudante é escolhido ao acaso e tem mais de $1,75m$. Qual a probabilidade de que seja homem?
8. Um certo programa pode ser usado com uma entre duas sub-rotinas A e B, dependendo do problema. A experiência tem mostrado que a sub-rotina A é usada 40% das vezes e B 60% das vezes. Se A é usada, existe 75% de chance de que o programa chegue a um resultado dentro do limite de tempo. Se B é usada, a chance é de 50%. Se o programa foi realizado dentro do limite de tempo, qual a probabilidade de que a sub-rotina A tenha sido a escolhida?