

Exercícios
Análise exploratória

A maioria dos exercícios foram extraídos de seções do livro: Magalhães, MN; Lima, ACP. Noções de Probabilidade e Estatística. São Paulo: EDUSP, 2008.

1. Classifique em verdadeiro ou falso as seguintes afirmações:
 - a) Estatística é um conjunto de técnicas destinadas a organizar um conjunto de valores numéricos.
 - b) Sempre que estivermos trabalhando com números, deveremos utilizar a Inferência Estatística.
 - c) A Estatística Descritiva fornece uma maneira adequada de tratar um conjunto de valores, numéricos ou não, com a finalidade de conhecermos o fenômeno de interesse.
 - d) Qualquer amostra representa, de forma adequada, uma população.
 - e) As técnicas estatísticas não são adequadas para casos que envolvam experimentos destrutivos, por exemplo, queima de equipamentos, destruição de corpos de prova, etc.
2. Para as situações descritas a seguir, identifique a população e a amostra correspondente. Discuta a validade do processo de inferência estatística para cada um dos casos.
 - a) Para avaliar a eficácia de uma campanha de vacinação no Estado de São Paulo, 200 mães de recém-nascidos, durante o primeiro semestre de um dado ano e em uma dada maternidade de São Paulo, foram entrevistadas a respeito da última vez em que vacinaram seus filhos.
 - b) Uma amostra de sangue foi retirada de um paciente com suspeita de anemia.
 - c) Para verificar a audiência de um programa de TV, 563 indivíduos foram entrevistados por telefone com relação ao canal em que estavam sintonizados.
 - d) A fim de avaliar a intenção de voto para presidente dos brasileiros, 122 pessoas foram entrevistadas em Brasília.
3. Discuta, para cada um dos casos abaixo, os cuidados que precisam ser tomados para garantir uma boa conclusão a partir da amostra.
 - a) Um grupo de crianças será escolhido para receber uma nova vacina contra meningite.
 - b) Sorteamos um certo número de donas de casa, para testar um novo sabão em pó.
 - c) Uma fábrica deseja saber se sua produção de biscoitos está com o sabor previsto.
 - d) Aceitação popular de um certo projeto do governo.
4. Classifique cada uma das variáveis abaixo em qualitativa (nominal / ordinal) ou quantitativa (discreta / contínua).
 - a) Ocorrência de hipertensão pré-natal em grávidas com mais de 35 anos (*sim* ou *não* são possíveis respostas para esta variável).
 - b) Intenção de voto para presidente (possíveis respostas são os nomes dos candidatos ou *não sei*).
 - c) Perda de peso de maratonistas na Corrida de São Silvestre, em gramas.
 - d) Intensidade da perda de peso de maratonistas na Corrida de São Silvestre (leve, moderada, forte).
 - e) Grau de satisfação da população brasileira com relação ao trabalho de seu presidente (valores de 0 a 5, com 0 indicando totalmente insatisfeito e 5 totalmente satisfeito).
5. Classifique os estudos descritos na sequência como experimentais ou observacionais.
 - a) Habitantes de um município foram consultados quanto às áreas mais carentes de investimentos por parte da administração pública;
 - b) Automóveis que circulam por uma rodovia foram selecionados inspecionados com o objetivo de analisar a emissão de poluentes;
 - c) Dois tipos de pintura foram aplicados a móveis de madeira com o objetivo de avaliar a aderência da tinta;
 - d) Estudantes dos cursos de Estatística, Agronomia e Enfermagem foram questionados e comparados quanto às expectativas profissionais;

- e) As capacidades respiratórias de funcionários fumantes e não fumantes foram verificadas numa pesquisa;
 - f) Numa plantação, três tipos de fertilizantes foram administrados a diferentes canteiros, e as consequentes produções foram registradas;
 - g) Automóveis de certo modelo foram analisados quanto ao desempenho registrado quando abastecidos com álcool e gasolina;
 - h) As preferências de consumidores de certo produto foram registradas quando apresentados a dois tipos de embalagens.
6. Relacione os tipos de amostragens com seus significados: Aleatória simples; estratificada; por conglomerado; sistemática.
- a) Os elementos da população estão ordenados e a retirada dos elementos é feita periodicamente.
 - b) A população já está dividida em diferentes grupos, portanto extrai-se uma amostra dos grupos selecionados, esperando que represente toda a população.
 - c) Considera a população homogênea. Cada elemento da população tem a mesma chance de ser escolhido.
 - d) Divide-se uma população heterogênea em grupos homogêneos e faz-se uma amostragem dentro de cada grupo.
7. Suponha que uma pesquisa de opinião pública deve ser realizada em um estado que tem duas grandes cidades e uma zona rural. Os elementos na população de interesse são todos os homens e mulheres do estado com idade acima de 21 anos. Que tipo de amostragem você sugere?
8. Comente sobre os seguintes planos amostrais, apontando suas incoerências, quando for o caso.
- a) Com a finalidade de estudar o perfil dos consumidores de um supermercado, observaram-se os consumidores que compareceram ao supermercado no primeiro sábado do mês.
 - b) Com a finalidade de estudar o perfil dos consumidores de um supermercado, fez-se a coleta de dados durante um mês, tomando a cada dia, um consumidor da fila de cada caixa do supermercado, variando sistematicamente o horário da coleta dos dados.
 - c) Para avaliar a qualidade dos itens que saem de uma linha de produção, observaram-se todos os itens das 14:00 às 14:30 horas.
 - d) Para avaliar a qualidade dos itens que saem de uma linha de produção, observou-se um item a cada meia hora, durante todo o dia.
 - e) Para estimar a porcentagem de empresas que investiram em novas tecnologias no último ano, enviou-se um questionário a todas as empresas. A amostra foi formada pelas empresas que responderam o questionário.
9. Em cada uma das situações abaixo, identifique o tipo de amostragem utilizado. Trata-se de um procedimento de amostragem probabilístico ou não probabilístico? Você teria alguma crítica ou sugestão de outro método de amostragem?
- a) Com o objetivo de analisar o perfil dos turistas que se destinam, por meio de excursões, a certa localidade turística, foi selecionada aleatoriamente uma amostra de ônibus de excursão, e então todos os turistas que viajavam nesses ônibus foram relacionados para compor a amostra de turistas.
 - b) Deseja-se investigar a proporção de torcedores dos times de certa localidade que tem o hábito de comparecer aos estádios. Para isso, foi feito um levantamento em que torcedores de cada time foram entrevistados até se obter uma amostra de tamanho desejado, com representação proporcional a dos torcedores na população.
 - c) Para se estimar o volume de madeira de árvores em uma reserva, as árvores foram marcadas com diferentes numerações, sorteando-se, de maneira completamente aleatória, as árvores a compor a amostra conforme as numerações atribuídas.
 - d) Num estudo sobre as condições estruturais de imóveis localizados em certa região, esses imóveis foram divididos conforme a idade da construção (inferior a 3 anos; entre 3 e 10 anos; ou superior a 10 anos), e uma amostra foi selecionada aleatoriamente em cada um desses grupos;
 - e) Para se estudar a satisfação de funcionários de uma empresa quanto às condições de trabalho, optou-se por realizar a amostragem em uma particular unidade, e os indivíduos a compor a amostra foram indicados pelo setor de recursos humanos, buscando compor uma amostra representativa da população de empregados.
 - f) Num estudo sobre o tempo de prateleira de produtos comercializados por um supermercado, foram selecionados os 10 itens dispostos mais à frente nas prateleiras, para cada um desses produtos, e as datas

de fabricação e validade foram registradas.

- g) Para se avaliar a disposição para venda por parte de proprietários de residências em uma importante avenida, selecionou-se, aleatoriamente, um imóvel da primeira quadra e, a partir dele, a cada dez imóveis o décimo também era relacionado para compor a amostra.
- h) Para se analisar a prevalência de lesões em corredores amadores, a cada indivíduo entrevistado pedia-se a indicação de algum conhecido que também praticasse corrida para compor a amostra.

10. Uma pesquisa tinha como objetivo verificar qual o meio de transporte mais comum entre os alunos de uma turma. As respostas foram:

Carro	Ônibus	Ônibus	Outro	Carro	A pé	Outro	Moto	Moto	Carro
Carro	Carro	Moto	Ônibus	Moto	Ônibus	Moto	Bicicleta	Ônibus	Moto
Outro	Carro	Moto	Carro	Carro	Ônibus	Carro	Ônibus	A pé	Ônibus

Construa uma tabela usando frequências absolutas e relativas (ou percentuais). Usando esta tabela esboce uma visualização adequada.

11. Diversas pesquisas de opinião fazem uso de um recurso chamado escala Likert. A ideia é que afirmativas são apresentadas e o respondente assinala o seu grau de concordância com cada afirmativa. Suponha que uma pesquisa de opinião sobre a qualidade dos restaurantes universitários está sendo conduzida e uma das afirmativas é “os produtos utilizados são de alta qualidade”. Um conjunto de respondentes tinha a possibilidade de assinalar um de cinco pontos: discordo totalmente (DT); discordo (D); indiferente (I); concordo (C) ou concordo totalmente (CT).

Considere o seguinte vetor de respostas:

C	D	I	C	DT	DT	C	CT	I	I	I	C	D	C	I
CT	DT	C	I	C	I	I	CT	C	C	I	C	I	DT	C

Construa uma tabela usando frequências absolutas, relativas (ou percentuais) e acumuladas. Usando esta tabela esboce uma visualização adequada.

12. Uma pesquisa com usuários de transporte coletivo na cidade de São Paulo indagou sobre os diferentes tipos nas suas locomoções diárias. Dentre ônibus, metrô e trem o número de diferentes meios de transporte foi o seguinte: 2, 3, 2, 1, 2, 1, 2, 1, 2, 3, 1, 1, 1, 2, 2, 3, 1, 1, 1, 2, 1, 1, 2, 2, 1, 2, 1, 2 e 3.
- a) Organize uma tabela de frequência.
 - b) Faça uma representação gráfica.
 - c) Admitindo que essa amostra represente bem o comportamento do usuário paulistano, você acha que a porcentagem dos usuários que utilizam mais de um tipo de transporte é grande?
13. Um novo medicamento para cicatrização está sendo testado e um experimento é feito para estudar o tempo (em dias) de completo fechamento em cortes provenientes de cirurgia. Uma amostra em trinta cobaias forneceu os valores: 15, 17, 16, 15, 17, 14, 17, 16, 16, 17, 15, 18, 14, 17, 15, 14, 15, 16, 17, 18, 18, 17, 15, 16, 14, 18, 18, 16, 15 e 14.
- a) Organize uma tabela de frequência.
 - b) Que porcentagem das observações estão abaixo de 16 dias?
 - c) Classifique como *rápida* as cicatrizações iguais ou inferiores a 15 dias e como *lenta* as demais. Faça um gráfico de setores indicando as porcentagens para cada classificação.
14. Um grupo de pedagogos estuda a influência da troca de escolas no desempenho de alunos do ensino fundamental. Como parte do levantamento realizado, foi anotado o número de escolas cursadas pelos alunos participantes do

estudo.

Escolas Cursadas	Frequência
1	46
2	57
3	21
4	15
5	4

- Qual é a porcentagem dos alunos que cursaram mais de uma escola?
- Construa o gráfico de barras.
- Classifique os alunos em dois grupos segundo a rotatividade: *alta* para alunos com mais de 2 escolas e *baixa* para os demais. Obtenha a tabela de frequência dessa variável.

15. Responda certo ou errado, justificando:

- Suponha duas amostras colhidas de uma mesma população, sendo uma delas de tamanho 100 e outra de tamanho 200. Então, a amostra de tamanho maior é mais representativa da população.
- Duas variáveis diferentes podem apresentar histogramas idênticos.

16. Um grupo de 84 estudantes do ensino médio foi submetido a um teste de matemática resultando em:

Nota	Frequência
[0; 2)	14
[2; 4)	28
[4; 6)	27
[6; 8)	11
[8; 10)	4

- Construa o histograma.
- Se a nota mínima para aprovação é 5, qual será a porcentagem de aprovação?

17. Alunos da Escola de Educação Física foram submetidos a um teste de resistência quanto ao número de quilômetros que conseguiram correr sem parar. Os dados estão apresentados a seguir.

Classes	Frequência
[0; 4)	438
[4; 8)	205
[8; 12)	125
[12; 16)	22
[16; 20)	9

- Qual é a variável em estudo?
- Construa o histograma.

18. Foram feitas medidas em operários da construção civil a respeito da taxa de hemoglobina no sangue em gramas por centímetros cúbicos, apresentando os seguintes resultados: 11.1, 12.2, 11.7, 12.5, 13.9, 12.3, 14.4, 13.6, 12.7, 12.6, 11.3, 11.7, 12.6, 13.4, 15.2, 13.2, 13.0, 16.9, 15.8, 14.7, 13.5, 12.7, 12.3, 13.5, 15.4, 16.3, 15.2, 12.3, 13.7 e 14.1.

- Organize os dados em classes de tamanho 1 a partir do 11.
- Construa o histograma.
- Taxas abaixo de 12 ou acima de 16 são consideradas alteradas e requerem acompanhamento médico. Obtenha a tabela de frequência da variável *Acompanhamento Médico* com duas opções: sim ou não.

19. Discuta quais medidas de posição seriam mais adequadas para os conjuntos de dados abaixo. Comente suas escolhas.

- Estão disponíveis dados mensais sobre a incidência de envenenamento por picada de cobra. Deseja-se planejar a compra mensal de antídoto.
- O número diário de usuários, entre 17 e 19 horas, de determinada linha de ônibus foi anotado. Pretende-se utilizar essa informação para dimensionar a frota em circulação.

- c) Um fabricante de baterias deseja divulgar a durabilidade do seu produto e coleta informação sobre a duração de 100 de suas baterias.
- d) Num voo internacional uma companhia serve dois tipos de pratos de jantar: peixe ou frango. Um banco de dados contém os pedidos feitos nos últimos 200 voos. Pretende-se planejar o número de cada tipo a ser colocado à disposição dos passageiros.
20. Vinte e cinco residências de um certo bairro foram sorteadas e visitadas por um entrevistador que, entre outras questões, perguntou sobre o número de televisores. Os dados foram os seguintes:

2, 2, 2, 3, 1, 2, 1, 1, 1, 1, 0, 1, 2, 2, 2, 3, 1, 1, 3, 1, 2, 1, 0 e 2.

Organize os dados numa tabela de frequências e determine as diversas medidas de posição.

21. Num experimento, 15 coelhos foram alimentados com uma nova ração e seu peso avaliado ao fim de um mês. Os dados referentes ao ganho de peso (em quilogramas) foram os seguintes:

1.5; 1.6; 2.3; 1.7; 1.5; 2.0; 1.5; 1.8; 2.1; 2.1; 1.9; 1.8; 1.7; 2.5 e 2.2.

- a) Utilizando os dados brutos, determine média, moda e mediana desse conjunto.
- b) Organize uma tabela de frequência com classes de amplitude 0.2 a partir de 1.5.
- c) Calcule, a partir da tabela de frequências e com o ponto médio como representante de cada classe, a média, a moda e a mediana. Comente as diferenças encontradas com o item (a).
- d) Se ao invés de 15, fossem 500 coelhos, qual seria o procedimento mais conveniente: o do item (a) ou o de (c)? Ter acesso a computador faz diferença?
22. Foram feitas medidas em operários da construção civil a respeito da taxa de hemoglobina no sangue em gramas por centímetros cúbicos, apresentando os seguintes resultados: 11.1, 12.2, 11.7, 12.5, 13.9, 12.3, 14.4, 13.6, 12.7, 12.6, 11.3, 11.7, 12.6, 13.4, 15.2, 13.2, 13.0, 16.9, 15.8, 14.7, 13.5, 12.7, 12.3, 13.5, 15.4, 16.3, 15.2, 12.3, 13.7 e 14.1.
Determine o terceiro quartil e a mediana.
23. A pulsação de 10 estudantes no início de uma prova de estatística foram as seguintes (em batimentos por minuto): 80, 91, 84, 86, 93, 88, 80, 89, 85 e 86. Calcule a média e a variância desse conjunto de dados.
24. Estudando uma nova técnica de sutura, foram contados os dias necessários para a completa cicatrização de determinada cirurgia. Os resultados de 25 pacientes foram os seguintes: 6, 8, 9, 7, 8, 6, 6, 7, 8, 9, 10, 7, 8, 10, 9, 9, 9, 7, 6, 5, 7, 7, 8, 10 e 11. Organize os dados numa tabela de frequência e calcule a média e a variância.
25. As notas finais de estatística para alunos de um curso de Administração foram as seguintes: 7, 5, 4, 5, 6, 3, 8, 4, 5, 4, 6, 4, 5, 6, 4, 6, 6, 3, 8, 4, 5, 4, 5, 5 e 6.
- a) Determine a mediana e a média.
- b) Separe o conjunto de dados em dois grupos denominados *aprovados*, com nota maior ou igual a 5, e *reprovados* para os demais. Compare a variância desses dois grupos.
26. Foram anotados os níveis de colesterol (em mg/100ml) para trinta pacientes de uma clínica cardíaca. As medidas se referem a homens entre 40 e 60 anos de idade que foram à clínica fazer um *check-up*.

Paciente	1	2	3	4	5	6	7	8	9	10
Colesterol	160	160	161	163	167	170	172	172	173	177

Paciente	11	12	13	14	15	16	17	18	19	20
Colesterol	178	181	181	182	185	186	194	197	199	203

Paciente	21	22	23	24	25	26	27	28	29	30
Colesterol	203	205	206	206	208	209	211	214	218	225

- a) Calcule a média, a moda e a variância a partir dos dados.
- b) Construa o *box-plot* dos dados.
- c) Existe algum paciente com nível de colesterol atípico em relação aos demais pacientes do grupo? Justifique sua resposta.

27. Um grupo de 84 estudantes do ensino médio foi submetido a um teste de matemática resultando em:

Nota	Frequência
1	14
3	28
5	27
7	11
9	4

Obtenha o gráfico de *box-plot*.

28. O setor de emergência de um Pronto Socorro Infantil anotou o *número de crianças atendidas* (C), *de médicos* (M) e *de auxiliares* (A) de plantão em 15 dias de atividade. Os dados são apresentados na tabela abaixo:

Dia	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
C	5	7	5	6	5	5	7	5	6	6	7	5	5	6	6
M	1	1	2	2	2	1	2	1	1	2	2	2	2	2	2
A	4	4	5	6	7	7	6	5	5	6	7	7	6	6	7

- Determine as tabelas de frequência marginais de C , M e A .
 - Obtenha a tabela de frequência conjunta entre (C, M) , (C, A) e (M, A) .
 - Represente a tabela de frequência conjunta de M e A , por meio de uma tabela de dupla entrada.
29. Na tabela a seguir, encontram-se os *conceitos de história* (H), *matemática* (M) e *física* (F) de alguns alunos do 3o. ano do ensino médio de uma escola:

Aluno	1	2	3	4	5	6	7	8	9	10	11	12
H	c	c	c	c	b	b	b	b	b	b	b	a
M	c	c	d	c	b	c	a	c	c	c	c	c
F	d	c	b	c	b	c	c	b	b	c	c	b

- Construa as tabelas de frequência conjunta para H e M e para H e F .
 - Calcule a proporção de alunos com “b” em física, dentre aqueles com pelo menos “c” em matemática.
30. Uma amostra de 220 clientes de uma clínica dentária foi selecionada. As variáveis tempo, em anos, decorridos desde a *última visita ao dentista* (V) e o *número de cáries encontradas* (C) são apresentados na tabela a seguir:

$V \setminus C$	0	1	2
1	18	16	10
2	34	45	38
3	12	16	31

Obtenha as tabelas de frequência marginais.

31. Para famílias de um certo bairro de São Paulo, apresentamos abaixo a tabela de frequência das variáveis *número de automóveis* (A) e *número de TVs* (T).

$A \setminus T$	0	1	2	total
0	110	235	120	465
1	51	122	178	351
2	15	74	162	261
total	176	441	460	1077

- Calcule as marginais de A e T .
 - Determine as médias de A e T .
32. Seja o seguinte quadro da concentração de albumina (g%) e do hormônio do crescimento (mm/ml) no plasma de carneiros.

Albumina	1.0	2.0	3.0	4.0	5.0	6.0	7.0
Hormônio	35.0	41.4	46.7	52.8	60.3	65.9	70.3

- a) Faça um gráfico adequado para verificar se há uma relação linear entre a concentração de albumina e hormônio do crescimento.
- b) Quantifique o grau de associação linear entre essas duas variáveis.
33. Um levantamento obtido, junto aos funcionários de um pequeno escritório, busca relacionar as variáveis: *anos de estudo* (X) e *número de diferentes emprego nos últimos 5 anos* (Y):

X	8	9	10	11	12
Y	4	2	1	2	1

O que você pode dizer sobre os dados fornecidos?

34. Um exame vestibular para uma faculdade tem 80 questões, sendo 40 de português e 40 de matemática. Para os 20 melhores classificados, apresentamos o número de acertos em cada disciplina, em ordem decrescente do total de pontos.

aluno	1	2	3	4	5	6	7	8	9	10
Português	35	35	34	32	31	30	26	26	24	23
Matemática	31	29	27	28	28	26	30	28	25	23

aluno	11	12	13	14	15	16	17	18	19	20
Português	23	12	11	20	17	12	14	20	8	10
Matemática	21	32	31	20	21	25	20	13	23	20

- a) Calcule as medidas de tendência central: média, mediana e moda para cada grupo.
- b) Calcule as medidas de variabilidade: variância, desvio-padrão, e coeficiente de variação para cada grupo.
- c) Calcule o resumo dos cinco números (mínimo, Q_1 , Q_2 , Q_3 e máximo) para cada grupo.
- d) Construa um gráfico de caixa e bigodes (*box-plot*) para cada grupo (em um mesmo gráfico para comparação).
- e) Com todos os resultados obtidos, descreva comparativamente estes dois grupos em termos de medidas de tendência central, variabilidade, amplitude e distribuição (simetria) dos dados.
- f) Você acha que os aprovados são melhores em português ou em matemática?

Respostas

1.
 - a) Falso. Estatística é mais do que organizar dados. Ela também se ocupa, entre outras coisas, de como coletar amostras e como extrapolar para toda a população os dados amostrados.
 - b) Falso. Se os dados numéricos se referem ao total da população não faz sentido utilizar a inferência estatística.
 - c) Verdadeiro. A Estatística Descritiva nos auxilia a explorar o conjunto de dados e é usualmente a primeira técnica a ser aplicada.
 - d) Falso. A amostra precisa ser coletada com cautela evitando vícios e distorções.
 - e) Falso. As técnicas estatísticas são especialmente úteis nos casos em que o objeto de estudo é danificado após sua experimentação, uma vez que minimiza o número de unidades que são investigadas.
2. Os enunciado dos itens desse exercício foram deixados propositadamente vagos de modo a permitir que os leitores discutam as alternativas de respostas. Abaixo apresentamos uma solução básica para cada item e ressaltamos que poderão haver outras soluções. O importante é identificar o conteúdo e a coerência da argumentação apresentada.
 - a) As crianças do Estado de São Paulo formam a população de interesse. As respostas das mães serão usadas para avaliar a eficácia da campanha de vacinação. A amostra de mães de recém nascidos de apenas uma maternidade não parece ser uma boa opção, entre outras, pelas seguintes razões: não temos informações sobre a vacinação nas várias regiões do estado nem nas várias camadas sociais; como vacinas não são dadas imediatamente em recém nascidos, precisaríamos que as mães entrevistadas tivessem outros filhos para poder dar a informação necessária. Extrapolar os resultados da amostra, nesse caso, parece ser inadequado.
 - b) O sangue do paciente é a população de interesse e por características biológicas ele é homogêneo em todo o corpo. Portanto, uma pequena quantidade, suficiente para os exames, poderá dar o panorama do que está acontecendo com o paciente. A inferência da amostra para toda a população é adequada.
 - c) A população seria o conjunto dos telespectadores da cidade. A amostra por telefone é fácil de fazer mas pode distorcer a verdade pois, dependendo da cidade, o telefone é um indicador de renda mais alta. Dessa forma, poderíamos estar viciando a amostra com um número excessivo de respostas da camada social mais elevada.
 - d) A população é constituída pelo conjunto de eleitores do país. A amostra, restrita a apenas uma população de cidade, não dá conta da diversidade de opiniões existentes no país, a menos que assuma que a população de Brasília represente bem o conjunto do país. Mesmo nesse caso, precisaríamos cuidados na escolha dos 122 eleitores. Por tudo isso, não parece ser adequado extrapolar o resultado dessa amostra para o país inteiro.
3. Esse problema pode ter várias soluções e portanto as respostas a seguir não devem ser interpretadas como as únicas corretas. Em todos os itens, o diálogo com os interessados na pesquisa dará mais consistência às soluções propostas.
 - a) É preciso identificar as características que poderiam influir no resultado da vacina. Por exemplo, será que idade, sexo e condição social, produzem diferentes resultados? Se esse é o caso, é conveniente garantir que a amostra contenha crianças com essas características em número proporcionalmente próximo ao que aparecem na população.
 - b) A eficiência do sabão em pó deve estar relacionada com o tipo e a condição da roupa a ser lavada, sortear as donas de casa num grande supermercado pode ser uma boa alternativa para mesclar essas características.
 - c) O grupo de provadores de biscoitos precisa ter registrado no seu paladar o sabor que precisa ser produzido. Da produção diária ou semanal é escolhida uma amostra de alguns biscoitos com o cuidado de alternar horários, máquinas produtoras, etc. Pode-se estabelecer um índice de semelhança com o sabor previsto, após ouvir a opinião dos provadores.
 - d) Dependendo do projeto do governo, pessoas com diferentes níveis de informação poderão ter opiniões diferentes. Classe social, renda ou sexo podem ser fatores que influenciam as respostas. Seria interessante

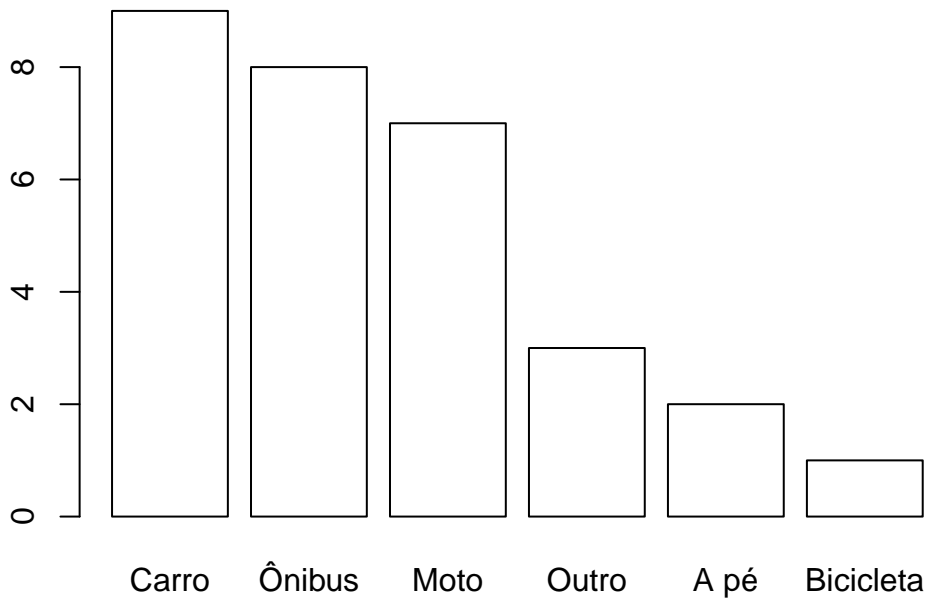
ter representantes, de cada um desses segmentos da população, na amostra a ser coletada.

-
4. a) Qualitativa nominal.
b) Qualitativa nominal.
c) Quantitativa contínua.
d) Qualitativa ordinal.
e) Qualitativa ordinal.
-
5. a) Estudo observacional, em que os habitantes são simplesmente questionados quanto às áreas mais carentes.
b) Estudo observacional, em que os automóveis têm suas emissões inspecionadas sem qualquer intervenção.
c) Estudo experimental, em que os móveis **são submetidos** a dois tipos diferentes de pintura, permitindo comparar a aderência da tinta nas duas condições.
d) Estudo observacional, em que os estudantes são simplesmente questionados quanto às suas expectativas profissionais.
e) Estudo observacional, em que os funcionários têm suas capacidades respiratórias, bem como suas condições quanto ao hábito de fumar, registradas.
f) Estudo experimental, em que cada canteiro foi tratado com um tipo de fertilizante, visando comparar as produções proporcionadas.
g) Estudo experimental, em que diferentes tipos de combustível (gasolina ou álcool) foram usados no abastecimento dos automóveis para fins de comparação de desempenho.
h) Estudo observacional, em que os consumidores foram apenas questionados quanto às suas preferências.
-
6. a) sistemática.
b) por conglomerado.
c) aleatória simples.
d) estratificada.
-
7. Note que podemos definir as classes de homens e mulheres como estratos da população. Portanto, a amostragem estratificada é adequada ao problema.
-
8. a) Pessoas que não fazem compras no primeiro sábado do mês não poderão fazer parte da amostra.
b) Considerando que pode haver uma periodicidade mensal de ida ao supermercado, esse plano se mostra satisfatório.
c) Qualquer tipo de anomalia no sistema fora deste horário não será verificada pela amostragem.
d) Uma amostra sistemática se mostra ideal para amostrar a fabricação de uma linha de produção, uma vez que a coleta é contínua e igualmente espaçada.
e) É comum esse tipo de pesquisa, já que muitas vezes é necessário contar com participantes voluntários. No entanto, a forma de amostragem poderá gerar algum tipo de viés devido a uma possível participação voluntária de um grupo muito específico.
-
9. a) Amostragem por conglomerados (probabilística), em que os ônibus são os conglomerados e os turistas as unidades amostrais.
b) Amostragem por cotas (não probabilística), em que as cotas correspondem às proporções de torcedores de cada time que devem compor a amostra.
c) Amostragem aleatória simples (probabilística), uma vez que as árvores amostradas são selecionadas de maneira completamente aleatória da população.
d) Amostragem estratificada (probabilística), em que as três faixas de idades dos imóveis compõem estratos, e os imóveis são selecionados aleatoriamente de cada um deles.

- e) Amostragem intencional ou por julgamento (não probabilística), dado que o setor de RH indica, sem qualquer seleção aleatória, um grupo de funcionários para compor uma amostra representativa da população.
- f) Amostragem por conveniência (não probabilística), pois a seleção dos itens se dá por acessibilidade, correspondendo às unidades posicionadas mais à frente nas prateleiras.
- g) Amostragem sistemática (probabilística), em que os imóveis são selecionados periodicamente ao longo da avenida, após a seleção aleatória do primeiro a compor a amostra.
- h) Amostragem bola de neve (não probabilística), dado que os indivíduos a compor a amostra são selecionados através de sucessivas indicações.

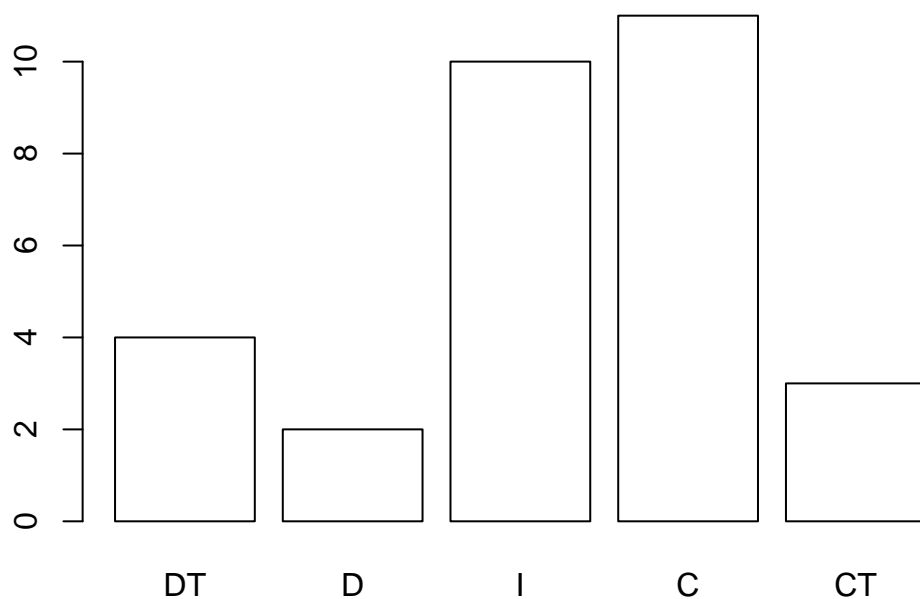
10.

Respostas	f_a	f_r
Carro	9	0.300
Ônibus	8	0.267
Moto	7	0.233
Outro	3	0.100
A pé	2	0.067
Bicicleta	1	0.033
Total	30	1.000



11.

Respostas	f_a	f_r	F_a	F_r
DT	4	0.133	4	0.133
D	2	0.067	6	0.2
I	10	0.333	16	0.533
C	11	0.367	27	0.9
CT	3	0.100	30	1
Total	30	1.000		



12. a)

Num. Transp.	1	2	3	total
frequência	14	12	4	30
freq. relativa	0.47	0.4	0.13	1

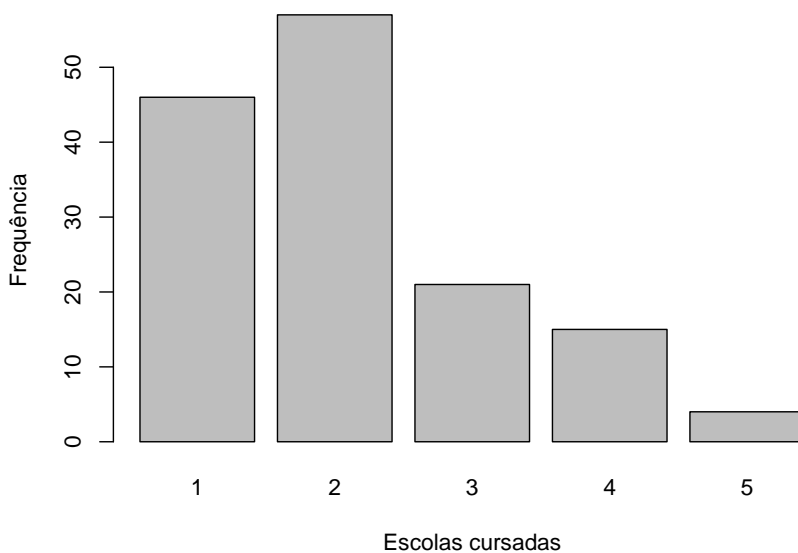
- b) Esses dados podem ser representados com um gráfico de setores ou um gráfico de barras.
 c) Sim, é mais de 50%.

13. a)

Cicatrização	14	15	16	17	18	total
frequência	5	7	6	7	5	30
freq. relativa	0.167	0.233	0.2	0.233	0.167	1

- b) 40%.
 c) Cicatrização rápida: 40% lenta: 60%.

-
14. a) 68%
 b)

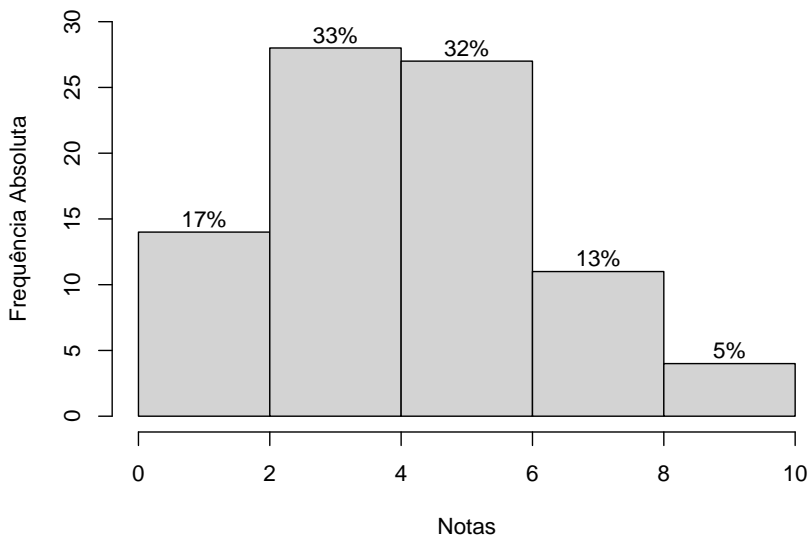


c)

Rotatividade	alta	baixa	total
frequência	40	103	143
freq. relativa	0.28	0.72	1

15. a) Errado. Nem sempre uma amostra maior irá representar melhor a população. Depende de como essa amostra foi coletada. Se as duas amostras forem coletadas com o mesmo cuidado usando o método de amostragem mais adequado, então a maior representará tão bem ou melhor do que a menor.
- b) Certo. Se as duas variáveis em questão forem tão correlacionadas que suas frequências sejam as mesmas.

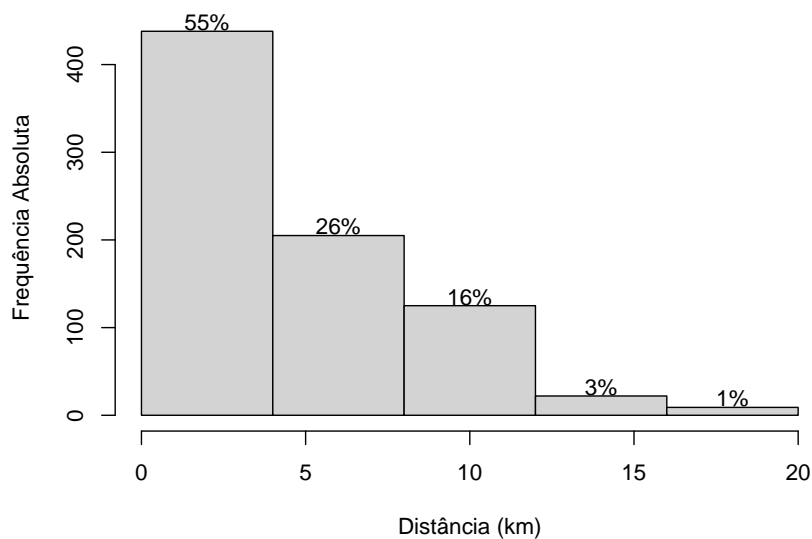
16. a)



- b) A nota mínima para aprovação é 5 e observando a tabela de frequências notamos que 5 está exatamente no centro do intervalo [4; 6] cuja frequência é 27. Para encontrar o percentual de estudantes aprovados vamos supor que as notas se distribuem de maneira uniforme dentro de cada intervalo e portanto esperamos que 50% das 27 notas do intervalo [4; 6] esteja no intervalo [5; 6], ou seja uma frequência de $27/2 = 13.5$ no intervalo [5; 6]. Assim, o percentual de aprovados será: $(50\% \text{ da frequência no intervalo } [4; 6] + \text{frequência no intervalo } [6; 8] + \text{frequência no intervalo } [8; 10]) / \text{número total de estudantes} \times 100 = \frac{((27 \times 0.5) + 11 + 4)}{14 + 28 + 27 + 11 + 4} \times 100 = 34\%$.

17.

- a) Quilômetros percorridos sem parar.
b) Construa o histograma.

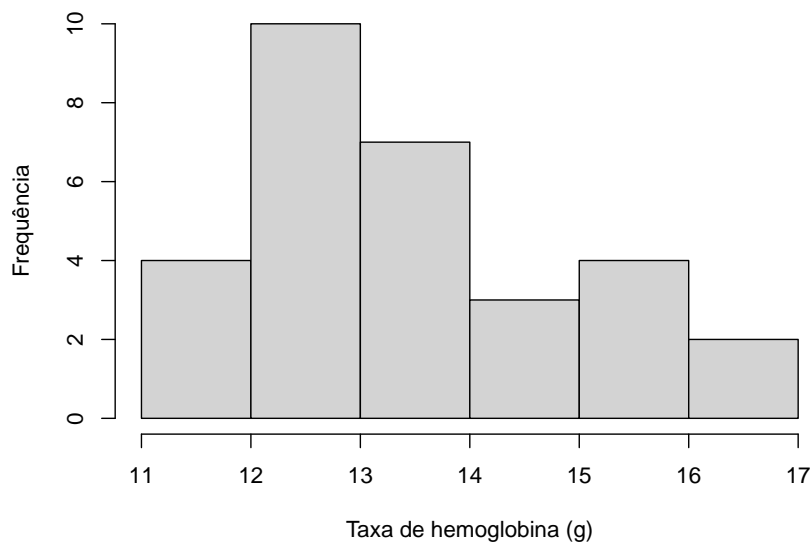


Classes	Frequência	Frequência Relativa	Frequência Acumulada	Frequência Relativa Acumulada
[0; 4)	438	0.55	438	0.55
[4; 8)	205	0.26	643	0.8
[8; 12)	125	0.16	768	0.96
[12; 16)	22	0.03	790	0.99
[16; 20)	9	0.01	799	1

18.

Taxa de hemoglobina (g)	[11; 12)	[12; 13)	[13; 14)	[14; 15)	[15; 16)	[16; 17)	total
a) frequência	4	9	8	3	4	2	30
freq. relativa	0.13	0.3	0.27	0.1	0.13	0.07	1

b)



c)

Acompanhamento médico	Sim	Não	total
frequência	6	24	30
freq. relativa	0.2	0.8	1

19. Este exercício é para possibilitar discussão entre os alunos sobre as várias formas de resumir a informação e portanto não existe uma resposta correta.

- O valor mediano deve ser suficiente em cerca de 50% dos meses. Se há meses com grande número de acidentes por questões sazonais como colheita ou enchentes, a média será afetada. É bom estar atento se a variabilidade dos dados não é grande.
- Para dimensionar a frota de ônibus podemos decidir do ponto de vista do usuário ou da empresa de ônibus. Do ponto de vista do usuário, pretende-se mais espaço sentado e ônibus não muito cheio; a empresa tentaria colocar o maior número de ônibus circulando. O número médio diário de passageiros nesse período pode ser dividido pelo número de assentos ou pela lotação total do ônibus dependendo de qual interesse pretende-se atender.
- A duração média é talvez a melhor medida a observar. Entretanto, se a moda for um valor alto (longa duração) pode ser útil utilizá-la como publicidade.
- Pode-se definir uma variável quantitativa indicando a proporção de quantos pedem frango (ou peixe) em cada voo. A moda dessa variável multiplicada pelo número total de passageiros no voo, poderia ser a quantidade de pratos colocados à disposição para tentar garantir o atendimento de todos os pedidos.

20. Tabela de frequência:

TVs	0	1	2	3	total
Freq.	2	10	10	3	25

$\bar{y} = 1.56$; $md = 13$ o. termo em ordem crescente = 2; $mo = 1$ ou 2.

21. a) Com os dados brutos $\bar{y} = 1.88$; $md = 1.8$ (observação que ocupa a 8a. posição em ordem crescente); $mo = 1.5$.

b) Tabela de frequência

Ganho	[1.5; 1.7)	[1.7; 1.9)	[1.9; 2.1)	[2.1; 2.3)	[2.3; 2.5]	total
Freq.	4	4	2	3	2	15

- Com a tabela de frequência no item anterior e usando o ponto médio da faixa obtemos: $\bar{y} = 1.933$; $md = 1.8$ e $mo = [1.5, 1.7); [1.7, 1.9)$. As diferenças entre as soluções de (a) e (c) não foram grandes nesse caso. A solução de (a) é mais exata.
- Os valores obtidos no item (a) possuem maior precisão, contudo, para um volume grande de dados, o cálculo fica muito trabalhoso para ser feito sem a ajuda de um computador. Notamos que a organização em classes é conveniente e os valores encontrados nas duas formas de calcular não diferem muito.

22. $Q_2 = 13.3$ e $Q_3 = 14.4$.

23. Utilizando as expressões para dados não agrupados, temos: $\bar{y} = 86.20$; $s^2 = 18.18$.

24. Tabela de frequência:

Dias para cicatrização	5	6	7	8	9	10	11	total
Freq.	1	4	6	5	5	3	1	25

Utilizando as expressões para dados não agrupados, temos: $\bar{y} = 7.88$; $s^2 = 2.36$.

25. a) $md = 5$ e $\bar{y} = 5.12$

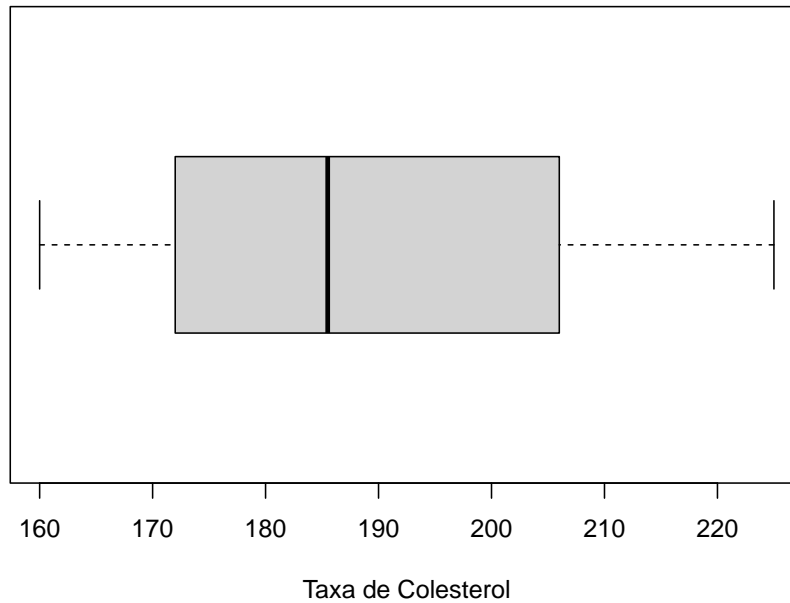
b) $s_{aprov.}^2 = 1.05 > s_{reprov.}^2 = 0.19$ mas $CV_{aprov.} = 17\%$ e $CV_{reprov.} = 12\%$ são próximos.

26. a) $\bar{y} = 188.87$; $md = 185.5$; $mo = 160, 172, 181, 203$ ou 206 e $s^2 = 369.29$.

b)

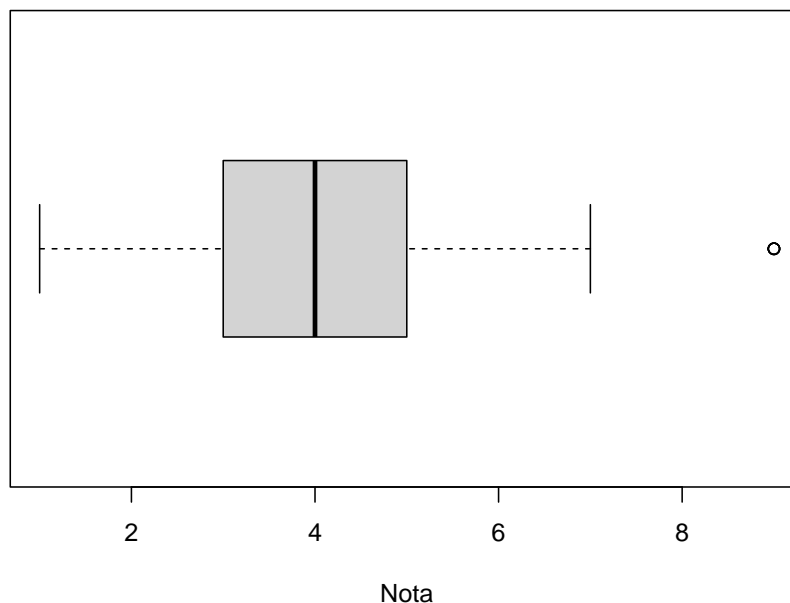
Boxplot	Min	Q_1	Q_2	Q_3	Máx
Colesterol	160	172	185.5	206	225

c) Não existe nenhum paciente com nível de colesterol além do limiar de 1,5 AIQ e portanto não há neste grupo nenhum paciente atípico em relação ao restante do grupo.



27. O boxplot construído a partir da tabela fornecida no enunciado é apresentado abaixo.

Box-plot: mínimo = 1, $Q_1 = 3$, $Q_2 = 4$, $Q_3 = 5$, máximo = 9.



Note que existem estudantes com nota atipicamente elevada quando comparados com os demais estudantes do grupo.

28. a)

C	n_i	f_i
5	7	0.47
6	5	0.33
7	3	0.20
total	15	1

M	n_i	f_i
1	5	0.33
2	10	0.67
total	15	1

A	n_i	f_i
4	2	0.13
5	3	0.20
6	5	0.33
7	5	0.34
total	15	1

- b) Para obter a tabela de frequência conjunta, contamos a frequência com que os pares de valores aparecem. Temos:

(C, M)	n_i
(5, 1)	3
(5, 2)	4
(6, 1)	1
(6, 2)	4
(7, 1)	1
(7, 2)	2
total	15

(C, A)	n_i
(5, 4)	1
(5, 5)	2
(5, 6)	1
(5, 7)	3
(6, 5)	1
(6, 6)	3
(6, 7)	1
(7, 4)	1
(7, 6)	1
(7, 7)	1
total	15

(M, A)	n_i
(1, 4)	2
(1, 5)	2
(1, 7)	1
(2, 5)	1
(2, 6)	5
(2, 7)	4
total	15

- c) As frequências conjuntas e marginais podem ser combinadas em uma única tabela de dupla entrada, conforme a tabela a seguir, para as variáveis M e A .

$M \setminus A$	4	5	6	7	total
1	2	2	0	1	5
2	0	1	5	4	10
total	2	3	5	5	15

29. a)

(H, M)	(c, d)	(c, c)	(b, c)	(b, b)	(b, a)	(a, c)	total
n_i	1	3	5	1	1	1	12

(H, F)	(c, d)	(c, c)	(c, b)	(b, c)	(b, b)	(a, b)	total
n_i	1	2	1	4	3	1	12

- b) 4/11.

30. Efetuando a soma de linha e de coluna obtemos as marginais:

V	n_i	f_i
1	44	0.20
2	117	0.53
3	59	0.27
total	220	1

C	n_i	f_i
0	64	0.29
1	77	0.35
2	79	0.36
total	220	1

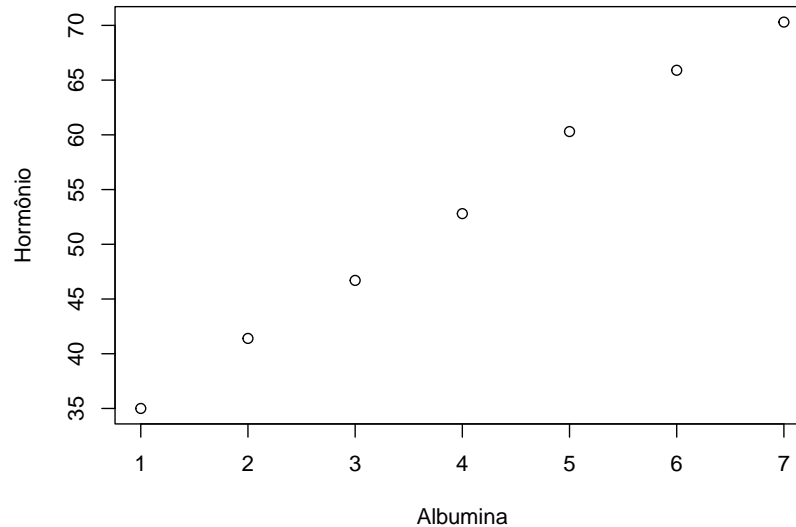
31. a) As frequências marginais são dadas pelas margens da tabela de dupla entrada. Portanto, temos:

A	n_i	f_i
0	465	0.43
1	351	0.33
2	261	0.24
total	1077	1

T	n_i	f_i
0	176	0.16
1	441	0.41
2	460	0.43
total	1077	1

b) $\bar{a} = 0.81$ e $\bar{t} = 1.27$.

32. a) Um gráfico adequado para visualizar a relação entre duas variáveis quantitativas é o diagrama de dispersão. Neste caso, há uma clara relação linear entre as duas variáveis.



- b) Para o cálculo do coeficiente de correlação de Pearson precisamos de $\sum_i (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2) = 168.5$, $\sum_i (y_{1i} - \bar{y}_1)^2 = 28$ e $\sum_i (y_{2i} - \bar{y}_2)^2 = 1017$. Assim o coeficiente é obtido por

$$r = \frac{\sum_i (y_{1i} - \bar{y}_1)(y_{2i} - \bar{y}_2)}{\sqrt{\sum_i (y_{1i} - \bar{y}_1)^2} \sqrt{\sum_i (y_{2i} - \bar{y}_2)^2}} = \frac{168.5}{\sqrt{28} \sqrt{1017}} = 0.9985$$

que indica uma forte associação positiva entre as variáveis.

33. Existe associação. Faça um gráfico dos pares (X, Y) e calcule $r = -0.775$.

34. a)

Medidas de centro	Port	Mat
\bar{y}	22.15	25.05
md	23	25.5
mo	c("12", "20", "23", "26", "35")	c("20", "28")

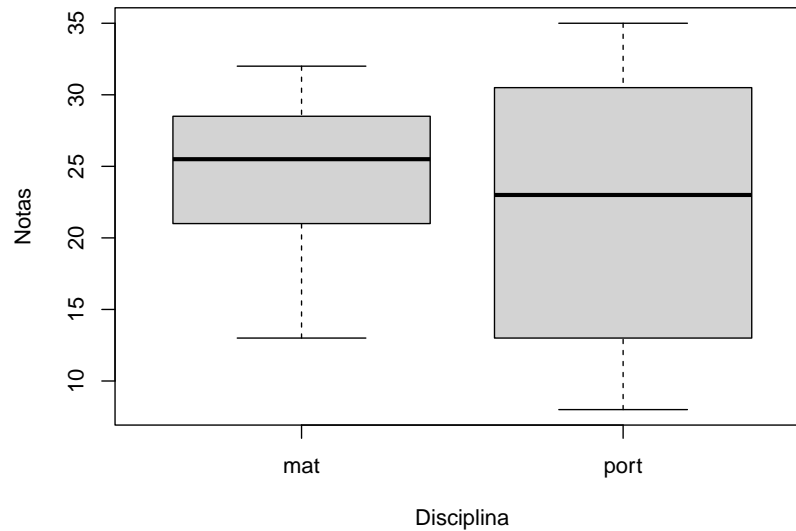
b)

Medidas de dispersão	Port	Mat
s^2	80.13	23.84
s	8.95	4.88
CV	40.4	19.5

c)

Medidas resumo	Port	Mat
Min	8	13
Q_1	13	21
Q_2	23	25.5
Q_3	30.5	28.5
Max	35	32

d)



- e) (*Exemplo de resposta*) Em média, o número de acertos em matemática ($\bar{y}_{mat} = 25.05$) foi maior do que o número de acertos em português ($\bar{y}_{port} = 22.15$). A diferença entre os valores médios e a mediana mostra que existe uma leve assimetria negativa (ou à esquerda) para os dois casos ($\bar{y} < md$), embora esta diferença seja mais pronunciada nas notas de português. A amplitude dos acertos em português foi de $A_{port} = 35 - 8 = 27$, maior do que a amplitude observada para o número de acertos em matemática, que foi de $A_{mat} = 32 - 13 = 19$. A variabilidade dos acertos em torno da média também foi maior para as notas de português, com variância de $s_{port}^2 = 80.13$ e desvio-padrão de $s_{port} = 8.95$. Já para a matemática, a variabilidade dos acertos em torno da média também foi menor, com $s_{mat}^2 = 23.84$ e desvio-padrão $s_{mat} = 4.88$. Resumindo estas informações sobre a variabilidade, nota-se que o coeficiente de variação para português foi de 40.4%, enquanto que para a matemática foi menor, com aproximadamente 19.5%. Através do resumo dos cinco números e do gráfico de caixa, percebe-se que 50% dos acertos foram entre 13 e 30.5 em português (diferença entre Q_1 e Q_3), e entre 21 e 28.5 em matemática, mostrando novamente a menor variabilidade observada para a matemática.

- f) Use os resultados obtidos nos itens anteriores para formular sua resposta.