

## Exercícios

### Estatística Descritiva

O conjunto de dados fornecido apresenta características de 22 pacientes com cisto no pâncreas:

<b>paciente</b>	<b>sexo</b>	<b>idade</b>	<b>tamanho do cisto (cm)</b>	<b>Localização do cisto no pâncreas</b>
1	F	49	6	cabeça
2	F	61	10	cabeça
3	M	34	8,2	cauda
4	F	73	3	colo
5	M	47	3,6	cabeça
6	M	58	10	colo
7	M	43	1	cabeça
8	M	71	1	cabeça
9	M	32	7	cauda
10	M	56	1	cabeça
11	M	61	6,6	corpo
12	F	49	4	cabeça
13	M	80	3,1	cauda
14	M	72	2,3	cabeça
15	M	47	10,5	cabeça
16	F	48	6,5	corpo
17	F	37	13	corpo
18	M	71	1	colo
19	M	74	7	cabeça
20	F	21	12	corpo
21	F	45	8,5	corpo
22	M	38	10	colo

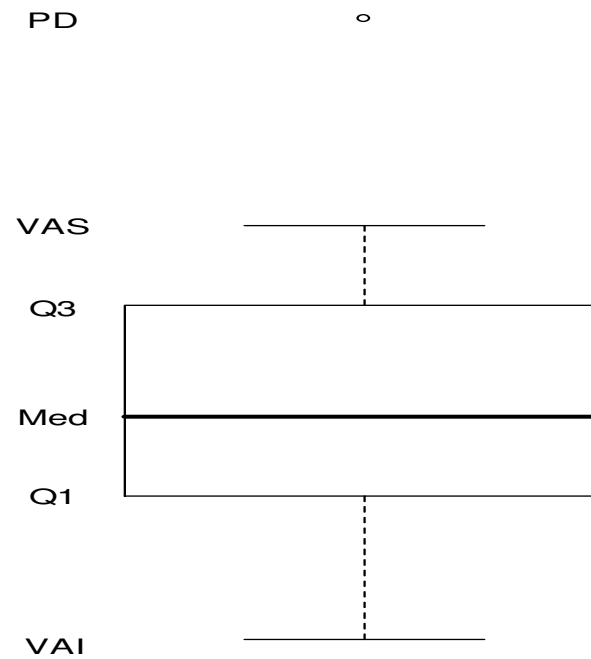
Pede-se:

- 1 – Construa uma tabela e um gráfico de distribuição de frequências para a localização do cisto no pâncreas dos pacientes;
- 2 – Construa uma tabela e um gráfico de distribuição de frequências conjuntas para as variáveis sexo e localização do cisto no pâncreas. Com base na tabela e gráfico obtidos, você suspeita que exista relação entre o sexo e a localização do cisto?
- 3 – Construa uma tabela e um gráfico de distribuição de frequências para os tamanhos dos cistos;
- 4 – São fornecidos gráficos dos tamanhos dos cistos segundo o sexo dos pacientes e suas localizações no pâncreas. Interprete os gráficos apresentados.
- 5 – Construa um gráfico que permita avaliar a relação entre a idade do paciente e o tamanho do cisto. Comente o resultado.
- 6 – Calcule as medidas de posição e dispersão estudadas para os tamanhos dos cistos. Depois, calcule-as novamente, para cada sexo. Você suspeita que exista relação entre o tamanho dos cistos e o sexo dos pacientes?

**Box-plots** – gráficos geralmente utilizados no estudo da distribuição dos valores de uma variável quantitativa para diferentes resultados de uma variável qualitativa.

Representação:

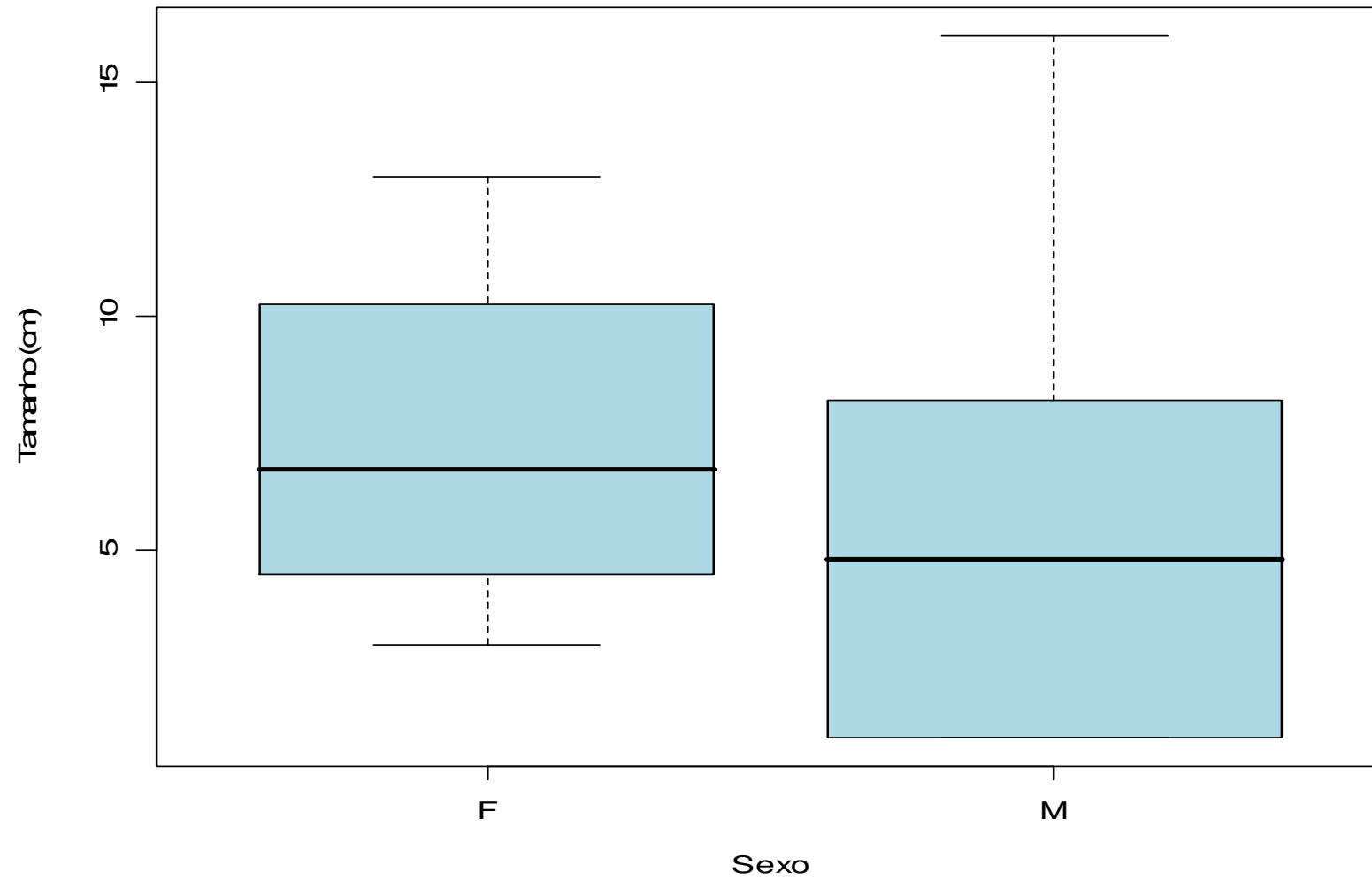
- ✓ Uma caixa, delimitada inferiormente pelo 1º quartil\* (Q1) e superiormente pelo 3º quartil\* (Q3), e segmentada em duas partes pela mediana (Med).
- ✓ A partir da extremidade superior da caixa, estende-se uma linha até o valor adjacente superior (VAS), valor da maior observação que esteja a uma distância inferior a uma vez e meia o tamanho da caixa ( $1,5 \cdot (Q3 - Q1)$ ) de Q3.
- ✓ O mesmo procedimento é adotado abaixo de Q1, tomando como valor adjacente inferior (VAI) a menor observação que esteja a uma distância inferior a uma vez e meia o tamanho da caixa.
- ✓ As observações afastadas da caixa por mais de  $1,5 \cdot (Q3 - Q1)$  são representadas por pontos e consideradas pontos discrepantes.



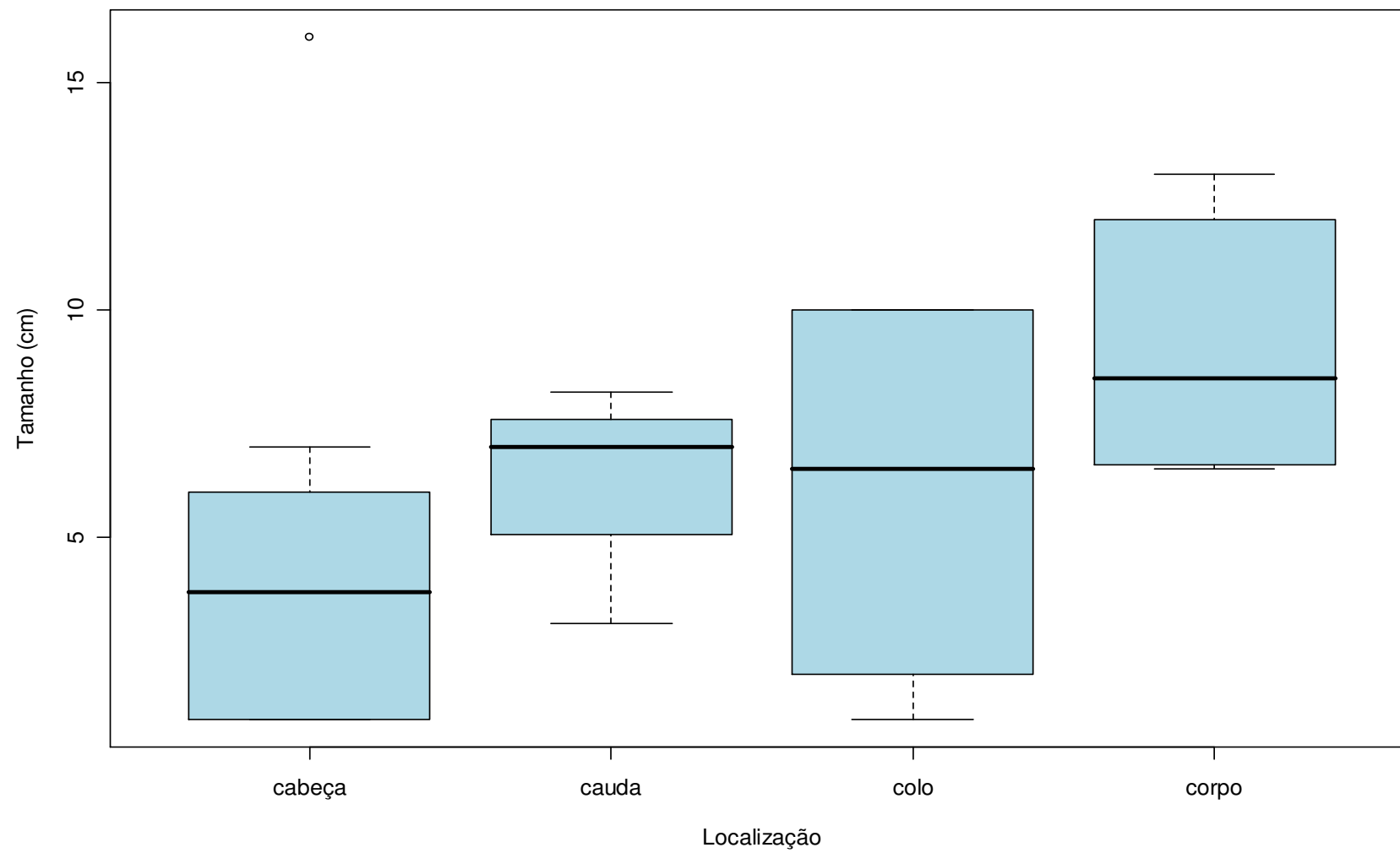
**Figura - Representação de um box-plot**

\*Os quartis são medidas de posição que dividem o conjunto ordenado de observações em quatro partes. A mediana é o valor que deixa 50% dos dados ordenados abaixo e 50% acima. O 1º quartil deixa 25% dos dados ordenados abaixo e 75% acima e o 3º quartil, 75% abaixo e 25% acima.

**Exemplo:** Box-plot do tamanho do cisto segundo o sexo do paciente.



**Exemplo:** Box-plot do tamanho segundo a localização do cisto.



### 1ª lista de exercícios

1. Os dados apresentados na tabela a seguir constituem uma amostra obtida num estudo sobre incidência de problemas no cotovelo entre jogadores de tênis.

**Tabela -** Dados de 20 membros de diversos clubes de tênis de Boston, EUA, que tiveram problema no cotovelo pelo menos uma vez. 1993.

Membro	Idade (anos)	Sexo**	Nº de vezes que teve o problema	Tamanho <sup>+</sup> da raquete*	Peso <sup>++</sup> da raquete*	Material <sup>-</sup> da armação da raquete*	Material <sup>-</sup> do trançado da raquete*
1	53	M	3	C	L	A	C
2	57	M	3	M	M	AL	N
3	43	M	1	C	M	G	N
4	35	F	2	C	L	FV	C
5	43	M	2	C	L	AL	C
6	31	M	1	C	L	G	C
7	36	M	1	C	M	M	C
8	36	F	1	M	L	AL	C
9	33	F	1	C	M	M	N
10	55	M	1	G	L	AL	N
11	35	F	1	M	L	AL	C
12	44	F	3	C	L	AL	N
13	45	F	8	C	L	FV	N
14	47	F	2	M	L	O	N
15	40	F	1	C	L	AL	C
16	41	F	1	C	L	M	N
17	40	F	8	C	L	O	C
18	40	F	3	M	L	AL	N
19	47	F	1	C	L	M	N
20	55	M	8	C	M	AL	N

**Fonte:** Rosner, B. *Fundamentals of biostatistics*. 4.ed. Duxbury Press, 1995. 682p. (Adaptação).

\* Raquete usada correntemente.

<sup>+</sup> C = Convencional.

M = Médio.

G = Grande.

-M = Madeira.

AL = Alumínio.

FV = Fibra de vidro.

G = Grafite.

A = Aço.

C= Composto.

<sup>++</sup> L = Leve.  
M = Médio.  
P = Pesado.

O = Outros.

<sup>-</sup> N = Nylon.  
C = Corda.

- Classifique as variáveis observadas em qualitativas (nominais ou ordinais) ou quantitativas (discretas ou contínuas);
- Construa uma tabela de distribuição de frequências para os tamanhos das raquetes utilizadas.
- Represente a tabela obtida no item 'b' por meio de um gráfico de colunas e de um gráfico de setores;

2. Para determinar as preferências de consumo na refeição principal, foi elaborado um estudo em indivíduos de ambos os sexos na cidade de Itajubá. A tabela seguinte resume os resultados obtidos:

Sexo	Tipo de refeição		
	Saladas	Carnes	Massas
Masculino	12	41	27
Feminino	35	15	30

- Construa uma tabela com as frequências relativas para cada sexo;
- Represente os dados através de um gráfico de colunas;
- Com base nos resultados obtidos nos itens anteriores, compare as preferências de consumo de homens e mulheres.

3. A taxa de mortalidade infantil corresponde ao número médio de mortes, dentre 1000 crianças nascidas vivas, antes de completarem um ano de vida. Os dados abaixo representam a Taxa de mortalidade infantil dos municípios da Microrregião Oeste Catarinense (1982) e foram extraídos da publicação *Municípios Catarinenses - Dados Básicos*, 1987, GAPLAN - SC, que utiliza dados levantados pelo IBGE.

32,3	62,2	10,3	22,0	13,1	9,9	18,3	33,0	20,0	22,7	27,2	11,9
36,4	23,5	18,0	22,6	20,3	38,3	32,9	29,9	29,7	39,2	25,4	19,6
	28,9	18,4	27,3	21,7	23,7	13,9	23,8	15,7	17,0	36,3	

- Agrupe convenientemente esses dados em classes e apresente a tabela de distribuição de frequências.
- Construa o histograma e o polígono de frequências.
- Calcule as medidas de posição e dispersão vistas em aula para as taxas de mortalidade.

4. Em uma investigação dos fatores de risco para as doenças cardiovasculares, os níveis séricos de cotinina (produto metabólico da nicotina) foram registrados para um grupo de fumantes e um grupo de não fumantes. As distribuições de frequências correspondentes são mostradas abaixo.

Nível de cotinina (mg/ml)	Fumantes	Não fumantes
0-13	78	3300
14-49	133	72
50-99	142	23
100-149	206	15
150-199	197	7
200-249	220	8
250-259	151	9
300-399	412	11
Total	1539	3445

- É correto comparar as distribuições dos níveis de cotinina para fumantes e não fumantes, com base nas frequências absolutas em cada intervalo? Por que?
  - Caso sua resposta para o item “a” seja negativa, construa uma nova tabela, em que as distribuições dos níveis de cotinina para fumantes e não fumantes possam ser comparadas.
  - Construa um gráfico com os polígono de frequências para fumantes e não fumantes.
  - Com base nos resultados obtidos nos itens “b” e “c”, o que você pode dizer sobre a distribuição dos níveis de cotinina registrados para cada grupo?
  - Para todos os indivíduos nesse estudo, o status do fumo é auto-registrado. Você acha que algum dos indivíduos pode estar mal classificado? Por que?
5. Em uma pesquisa sobre a concentração de minerais no leite materno, foram coletados no período de 1984 a 1985, dados de 55 mães do Hospital Maternidade Odete Valadares em Belo Horizonte. As mães foram divididas em dois grupos, segundo o período de lactação: colostro e leite maduro. Os minerais considerados foram cálcio, cobre, magnésio e zinco. Os dados a seguir referem-se ao cálcio e ao zinco.

Cálcio ( $\mu g / mL$ de leite) – Grupo colostro									
113	145	163	163	167	181	221	225	231	241
254	256	275	296	303	311	312	313	323	325
334	344	372	375	437					
Cálcio ( $\mu g / mL$ de leite) – Grupo maduro									
159	175	181	188	200	206	213	214	217	231
238	238	242	244	256	259	260	263	264	275



277	279	281	293	303	314	344	394		
Zinco ( $\mu\text{g} / \text{mL}$ de leite) – Grupo colostro									
1,07	1,20	1,30	3,13	3,20	3,70	4,40	4,57	5,20	5,82
6,07	6,13	6,50	6,82	6,82	6,90	7,23	7,42	7,43	7,43
7,45	8,25	8,40	8,77	9,54					
Zinco ( $\mu\text{g} / \text{mL}$ de leite) – Grupo maduro									
0,52	0,60	0,78	0,86	1,08	1,23	1,28	1,31	1,40	1,55
1,55	1,60	1,92	2,15	2,27	2,40	2,41	2,52	2,57	2,60
2,63	2,92	3,02	3,05	3,57	3,88	5,50	7,38		

Para as concentrações de cálcio e zinco em cada grupo:

- Calcule medidas de tendência central e de variabilidade, o coeficiente de variação e os quartis. Apresente essas estatísticas em uma tabela contendo inclusive o título. Comente os resultados.
- Construa boxplots e verifique a existência de outliers.
- Com base nos resultados dos itens anteriores, compare as concentrações dos minerais nos dois grupos.

### Exercícios de Bioestatística

1- O nível de colesterol no sangue é uma variável com distribuição Normal, de média  $\mu$  desconhecida e desvio padrão  $\sigma = 60\text{mg} / 100\text{ml}$ .

- Qual deve ser o valor de  $n$  para que apenas 10% das médias amostrais excedam a média populacional em 10 unidades ou mais?
- Teste a hipótese de que  $\mu = 260$ , contra a alternativa de que  $\mu > 260$  com base numa amostra de 50 pacientes, em que se observou uma média amostral  $\bar{x} = 268$ . Utilize um nível de significância de 5%.
- Construa um intervalo com 95% de confiança (bilateral) para o nível médio populacional de colesterol.
- Qual deve ser o tamanho amostral para que um intervalo de 99% de confiança para  $\mu$  tenha um comprimento de 30 unidades?

2- A fim de acelerar o tempo que um analgésico leva para penetrar na corrente sanguínea, um químico analista acrescentou certo componente à fórmula original, que acusava um tempo médio de 43 minutos. Em 36 observações com a nova fórmula, obteve-se um tempo médio de 42 minutos, com desvio padrão de 6 minutos.

- O que podemos concluir, ao nível de 5% de significância, sobre a eficiência do novo componente?
- Qual seria a resposta ao nível de 1%?
- Que tipo de erro pode ser cometido?
- Estime o tempo médio que o analgésico leva para penetrar na corrente sanguínea com um nível de 95% de confiança.

3- Entre milhares de casos de pneumonia não tratada com sulfa, a porcentagem que desenvolveu complicações foi de 10%. Com o intuito de saber se o emprego da sulfa diminuiria essa porcentagem, 120 casos de pneumonia foram tratados com sulfapiridina e destes, 6 apresentaram complicações.

- a) Teste a hipótese de que a proporção de casos com complicações dentre os pacientes tratados com sulfa é significativamente menor do que os não tratados, com base no nível descritivo do teste. Comente o resultado.
- b) Construa um intervalo com 95% de confiança para a porcentagem de complicações para doentes tratados com sulfa.

4- Num estudo sobre o metabolismo do citrato no fígado foram tomadas amostras de sangue da veia hepática de dez indivíduos normais e amostras de sangue arterial de outros dez indivíduos normais, obtendo-se as seguintes determinações de citrato em cada amostra (em mg/mL):

Sangue da veia hepática	20,2	24,6	18,3	19,0	29,5	12,6	18,2	30,8	22,2	25,4
Sangue arterial	26,4	32,2	37,8	25,0	28,4	26,2	31,3	35,0	29,7	27,4

- a) Forneça uma estimativa pontual para a diferença das médias de citrato no sangue da arterial em relação ao da veia hepática;
- b) Verifique se há diferença significativa no sentido de um maior conteúdo médio de citrato no sangue arterial em relação ao sangue da veia hepática, ao nível de significância de 5%.
- c) Calcule um intervalo com 95% de confiança para a diferença dos níveis médios de nitrato para o sangue arterial em relação ao da veia hepática.

5- Dois candidatos (A e B) a um emprego foram submetidos a um conjunto de oito questões, sendo anotados os tempos que cada um gastou na solução. Podemos, ao nível de 5% de significância, concluir que B é mais rápido que A, em termos do tempo médio gasto para resolver questões do tipo daquelas formuladas?

Questão	1	2	3	4	5	6	7	8
Tempo de A	11	8	15	2	7	18	9	10
Tempo de B	5	7	13	6	4	10	3	12

## Probabilidade

**Exercício 1**– Uma universidade tem 10 mil alunos dos quais 4 mil são esportistas. Temos ainda que 500 cursam medicina, 700 cursam biologia, 100 são esportistas e da medicina diurno e 200 são esportistas e da biologia. Um aluno é escolhido ao acaso e pergunta-se a probabilidade de:

- a. Ser esportista.
- b. Ser esportista e aluno da medicina.
- c. Ser esportista, dado que o aluno sorteado cursa medicina.

- d. Cursar medicina, dado que o aluno sorteado é esportista.
- e. Não ser esportista nem cursar biologia ou medicina.

**Exercício 2** – Um médico desconfia que um paciente tem tumor no abdômen, pois isto ocorreu em 70% dos casos similares que tratou. Se o paciente de fato tiver o tumor, o exame ultra-som o detectará com probabilidade 0,9. Entretanto, se ele não tiver o tumor, o exame pode, erroneamente, indicar que tem com probabilidade 0,1.

- a. Qual a probabilidade de o exame detectar tumor?
- b. Se o exame detectar o tumor, qual a probabilidade de o paciente tê-lo de fato?

**Exercício 3** – Uma certa doença pode ser curada através de procedimentos cirúrgicos em 80% dos casos. Dentre os que têm essa doença, sorteamos 15 pacientes que serão submetidos à cirurgia. Fazendo alguma suposição adicional que julgar necessária, responda:

- a. Qual a probabilidade de todos serem curados?
- b. Qual a probabilidade de ao menos dois não serem curados?
- c. Qual o número esperado de pacientes curados? Qual o desvio padrão?

**Exercício 4** – O número de crianças que dão entrada no setor de emergências de um hospital segue o modelo de Poisson, com taxa de chegada  $\lambda = 3$  pacientes por hora.

- a. Qual a probabilidade de, numa hora qualquer, uma criança dar entrada no setor de emergência?
- b. Qual a probabilidade de no máximo uma criança dar entrada no setor de emergências?

**Exercício 5** - Uma clínica de emagrecimento recebe pacientes adultos com peso seguindo uma distribuição normal de média 130Kg e desvio padrão 20Kg. Qual a probabilidade de um paciente escolhido ao caso pesar:

- a. Menos de 140Kg;
- b. Entre 125 e 135Kg;
- c. Mais de 150Kg.
- d. Para efeito de determinar o tratamento mais adequado, os 25% pacientes de menor peso são classificados como “magros”, enquanto os 25% de maior peso de “obesos”. Determine os valores que delimitam cada uma destas classificações.

- 1- Considerando que a probabilidade de nascer uma criança do sexo feminino é de 47%, qual a probabilidade de que em uma família com seis filhos pelo menos cinco sejam do sexo feminino?
- 2- Suponha que o processo de esterilização para um experimento biológico compreenda  $n$  estágios independentes, cada um com probabilidade  $p$  de sucesso. Se uma falha em qualquer dos estágios ocasiona contaminação, qual a probabilidade desta acontecer considerando  $n=10$  e  $p=0,99$ .
- 3- Um contador eletrônico de bactérias registra, em média, cinco bactérias por  $cm^3$  de um líquido. Admitindo-se que esta variável tenha distribuição de Poisson,
  - a. Qual é o desvio padrão do número de bactérias por  $cm^3$ ?
  - b. Encontre a probabilidade de que pelo menos duas bactérias ocorram num volume de líquido de  $1\text{ cm}^3$ .
- 4- A experiência mostra que  $2/3$  dos pacientes que têm certa doença recuperam-se sob um tratamento padrão. Uma nova droga promissora será administrada a um grupo de 12 pacientes que têm a doença. Se a clínica decidir aceitar a nova droga como o melhor tratamento apenas se no máximo um desses pacientes não se recuperar, qual a probabilidade de que esta seja desacreditada, ainda que a taxa de recuperação seja igual a 0,8?
- 5- É sabido que para adultos do sexo masculino, gozando de boa saúde, em uma certa população, a temperatura corporal segue distribuição normal com média de 36,8 graus e desvio padrão de 0,15 graus.
  - a. Qual a probabilidade de um adulto selecionado ao acaso dessa população apresentar temperatura entre 36,8 e 37,2 graus?
  - b. Se considerarmos 1000 dessas pessoas, quantas se esperariam com temperatura entre 36,8 e 37,2 graus?
  - c. Em qual intervalo de temperaturas estão 98% dos adultos masculinos sadios dessa população?
- 6- Calcule a mediana e os percentis de ordem 25 e 75 para uma variável com distribuição normal com média  $\mu$  e desvio padrão  $\sigma$ .
- 7- Os seguintes valores representam os tempos gastos, em segundos, por pessoas com doença cardíaca para a realização de um teste simples de esforço.

289	203	359	243	232	210
251	246	224	239	220	211

- a) Calcule a média, a mediana, o desvio padrão e os percentis de ordem 25 e 75.

- b) Com base nos resultados obtidos no exercício anterior, você diria que esses dados de fato foram gerados de uma distribuição normal? Justifique sua resposta.
- 8- O teste de conjuntos mede a função mental. É um teste rápido e simples e requer que um indivíduo se recorde de itens de quatro categorias comuns. O teste foi aplicado em 65 voluntários idosos sem disfunção mental e os resultados apresentados disfunção mental e os resultados apresentados na tabela a seguir. A média, desvio padrão e a mediana foram:  $\bar{x} = 30,63$  ;  $s = 7,67$  ;  $md = 32$ .

4	11	12	15	15	16	20	21	22	22	23	24	24
25	26	26	26	26	27	27	28	28	29	29	29	30
30	30	30	31	32	32	32	32	33	33	34	34	35
35	35	36	36	37	37	37	37	37	37	37	38	38
39	39	39	39	39	39	39	40	40	40	40	40	40

- a) Calcule o limite inferior de referência, que seja superado aproximadamente por 85% dos idosos testados. Use os métodos da curva normal e dos percentis;
- b) Calcule e interprete a especificidade deste teste;
- c) Neste caso, poderíamos calcular a sensibilidade? Por que?
- d) Se calculássemos o valor de referencia inferior baseado em 95%, ao invés de 85%, o que ocorreria com a especificidade? Por que?

### Exercícios – Estimação

1 - Alguns pesquisadores pensam que a vitamina C pode ser útil para reduzir os depósitos formadores do colesterol situados na parte inferior das paredes arteriais, e, por conseguinte, pensam que a vitamina C concorre também para a redução da possibilidade de ataques cardíacos. O nível de colesterol de cada uma de 36 pessoas com nível de colesterol acima do normal foi anotado antes e após um período de regime de 1 mês, regime este que obrigou cada pessoa a ingerir 500 mg de vitamina C por dia. Os dados obtidos mostraram 64,3 mg por 100 ml para a queda média do nível de colesterol. Suponha conhecido o desvio padrão populacional, igual a 18,9mg por 100 ml.

- a) Estime a queda média do nível de colesterol por pessoa, utilizando um intervalo de 95% de confiança.
- b) Qual o erro associado à estimativa?
- c) Qual o tamanho de amostra necessário para reduzirmos o erro para 3mg por 100ml?

2 – Os dados apresentados na sequência são tamanhos (em dias) dos ciclos menstruais em uma amostra aleatória de 15 mulheres.

26	28	29	33	29	26	30	31
28	27	29	30	32	28	29	

- Forneça uma estimativa pontual para o tamanho médio do ciclo menstrual;
- Teste a hipótese de que o tamanho médio do ciclo menstrual é inferior a 30 dias. Avalie quais seriam as conclusões aos níveis de 10, 5 e 1% de significância.
- Estime o tamanho médio do ciclo menstrual por meio de um intervalo com 95% de confiança. Qual a margem de erro de sua estimativa?

3- Para a população masculina de meia idade que mais tarde desenvolve a *diabetes mellitus*, a distribuição dos índices básicos de massa corpórea é aproximadamente normal, com média  $\mu$  e desvio padrão  $\sigma$  desconhecidos. Uma amostra de 58 homens selecionados desse grupo tem média  $\bar{x} = 27 \text{ kg} / \text{m}^2$  e desvio padrão  $s = 2,7 \text{ kg} / \text{m}^2$ .

- Teste a hipótese de que o índice médio dessa população é superior a  $25 \text{ kg} / \text{m}^2$ , índice máximo associado a pessoas saudáveis. Considere um nível de significância de 1%;
- Estime o índice médio de massa corpórea por meio de um intervalo com 99% de confiança.

4 - Uma amostra de 400 pessoas indicou que 240 foram classificadas, de acordo com sua capacidade de visão, como enxergando melhor com o olho direito.

- Formule adequadamente as hipóteses para avaliar se mais da metade da população enxerga melhor
- Estime o índice médio de massa corpórea por meio de um intervalo com 99% de confiança.

- Estime a fração de toda a população composta por pessoas que enxergam melhor com o olho direito. Use um coeficiente de confiança de 95%.
- Com base no intervalo resultante, você afastaria a possibilidade de que 50% das pessoas enxergam melhor com o olho direito? Justifique.

**Mais exercícios**

Os dados (fictícios) relativos aos exercícios desta lista encontram-se em um arquivo separado, também enviado por e-mail, e são específicos para cada uma das duplas. Considere, para cada um dos exercícios, **SOMENTE** os conjuntos de dados identificados com o número de sua dupla. A resolução dos exercícios deve ser descrita de maneira detalhada, com todos os seus cálculos, e os resultados obtidos devem ser devidamente adequados.

Exercício 1 - Deseja-se saber se mulheres e homens diferem quanto à preferência pela medicina tradicional ou alternativa para o tratamento de problemas médicos relacionados ao stress. Com tal finalidade, uma amostra de 200 indivíduos foi selecionada de maneira aleatória, tendo suas preferências registradas. Teste a hipótese de associação entre o sexo e o tipo de tratamento preferido. Apresente suas conclusões aos níveis de 10, 5 e 1% de significância.

Exercício 2 – Um experimento inteiramente casualizado é executado com o objetivo de comparar cinco medicamentos analgésicos segundo o tempo médio que cada um leva para aliviar dores de cabeça. Trinta pacientes participantes do experimento foram medicados com um dos cinco medicamentos sob estudo, registrando-se o tempo decorrido desde o momento em que foram medicados até o alívio da dor. Considere que os resultados do experimento são aqueles identificados pelo número de sua dupla.

- Calcule a média e a variância para cada um dos cinco medicamentos;
- Formule as hipóteses para um teste de igualdade de médias;
- Teste as hipóteses formuladas no item anterior considerando todos os níveis de significância para os quais você dispõe da tabela. O que você conclui?
- Caso você tenha detectado diferença entre as médias, ao nível de significância de 5%, execute o teste de comparações múltiplas de Tukey, também ao nível de 5%, e represente os resultados convenientemente.

Exercício 3 - Deseja-se estudar a correlação entre o desempenho dos alunos de Medicina nas disciplinas de Química e Estatística. Quinze alunos foram selecionados aleatoriamente e suas notas finais nas duas disciplinas registradas.

- Construa um gráfico de dispersão para esses dados;
- Há alguma evidência de uma relação linear entre as notas em Química e Estatística?
- Calcule o coeficiente de correlação de Pearson;
- Aos níveis de 10, 5 e 1% de significância, teste a hipótese de que a correlação é nula contra a hipótese alternativa de que a correlação é positiva;
- Construa um intervalo de 95% de confiança para a correlação populacional.

**Análise de variância e  
teste de comparações múltiplas**

**Exercício** - Os dados obtidos num experimento inteiramente ao acaso estão apresentados na tabela abaixo.

	Tratamento				
	A	B	C	D	E
	12	11	8	15	16
	13	8	11	17	17
	10	7	13	17	19
	13	9	12	17	16
	13	9	12	14	16
	11	10	10	16	18
$\sum x$	72	54	66	96	102
$\sum x^2$	872	496	742	1544	1742

- e. Calcule a média e a variância para cada um dos cinco tratamentos;
- f. Formule as hipóteses para um teste de igualdade de médias;
- g. Teste as hipóteses formuladas no item anterior considerando todos os níveis de significância para os quais você dispõe da tabela. O que você conclui?
- h. Uma das pressuposições para a validade dos resultados da análise de variância é que as variâncias dos tratamentos não difiram significativamente. Utilize a regra do quociente de variâncias ( $s_{\max}^2 / s_{\min}^2$ ) para avaliar a validade dos resultados.
- i. Caso você tenha detectado diferença entre as médias, ao nível de significância de 5%, execute o teste de comparações múltiplas de Tukey, também ao nível de 5%, e represente os resultados convenientemente.



Deseja-se estudar a correlação entre o desempenho dos alunos de Medicina nas disciplinas de Química e Estatística. Quinze alunos foram selecionados aleatoriamente e suas notas finais nas duas disciplinas registradas.

## PROVAS

**1-** Ensaios clínicos são experimentos utilizados nas Ciências Médicas com o objetivo de comparar tratamentos ou opções terapêuticas. O delineamento de estudos deste tipo requer que alguns requisitos básicos sejam satisfeitos, a fim de que os resultados produzidos sejam efetivamente comparáveis. As questões listadas na sequência versam a respeito de tais requisitos e outras características inerentes a ensaios clínicos.

- (1 ponto) Segundo o princípio da aleatorização, os tratamentos (ou opções terapêuticas) devem ser atribuídos às unidades experimentais (pessoas, cobaias,...) de maneira aleatória. Justifique o propósito deste requisito.
- (1 ponto) Fatores externos (outros, que não propriamente o tratamento) devem ser controlados no planejamento do experimento. Discuta a necessidade deste controle. Ilustre essa necessidade por meio de um exemplo.
- (1 ponto) É comum a execução de ensaios cegos (em que o paciente não sabe a medicação que recebe) ou até mesmo duplo-cegos (em que ambos médico e paciente desconhecem o tratamento aplicado). Justifique a razão de se executar ensaios cegos ou duplo-cegos.

**2-** Os dados apresentados na sequência mostram os resultados de colesterol total ( $mg/100ml$ ) em dois grupos de pacientes. O grupo T é formado por 10 pacientes submetidos a um controle alimentar, enquanto o grupo C é constituído por 10 pacientes sem nenhum tipo de tratamento ou controle.

Grupo C	222	247	264	291	294	299	331	332	421	454
Grupo T	195	203	211	223	232	240	241	242	258	267

Dicas: Grupo C:  $\sum x = 3155, \sum x^2 = 1043649$ ; Grupo T:  $\sum x = 2312, \sum x^2 = 539406$ ;

- (1 ponto) Calcule medidas de tendência central (média e mediana) e de dispersão (desvio padrão) para os níveis de colesterol em cada um dos grupos.
- (0,5 ponto) Construa um box-plot para os níveis de colesterol observados em cada um dos grupos. Represente-os numa única figura.
- (1 ponto) Com base nos resultados obtidos nos itens anteriores compare os dois grupos.

**3-** Os dados abaixo são tomados de um estudo que investiga o uso de uma técnica chamada ventriculografia radionuclídica como teste de diagnóstico para se detectar doença da artéria coronária.

Teste	Doença		Total
	Presente	Ausente	
Positivo	302	80	382
Negativo	179	372	551
Total	481	452	933

- a) (1 ponto) Qual a sensibilidade da ventriculografia radionuclídica nesse estudo? Qual a sua especificidade?
- b) (1 ponto) Para uma população cuja prevalência da doença coronária é de 0,10, calcule a probabilidade de que um indivíduo tenha a doença, dado que ele apresentou resultado positivo usando a ventriculografia radionuclídica?
- c) (1 ponto) Qual o valor preditivo de um teste negativo?

**4-** (1,5 pontos) Foram coletadas as frequências cardíacas de 50 jovens saudáveis. A frequência cardíaca média obtida foi igual a 80 b.p.m., com desvio padrão de 10 b.p.m. Determine uma faixa de referencia com cobertura de 95% usando o método da curva normal. O que aconteceria com a faixa se fixássemos uma cobertura de 99%? Por que?

**Boa Prova!**

1. Os valores abaixo representam os níveis de chumbo ( $x$ , expressos em  $\mu g / dl$ ) no sangue de um grupo de 20 trabalhadores.

15	25	32	38	42
45	48	50	52	53
54	57	60	63	66
72	78	84	92	100

- a) Agrupe os dados em cinco intervalos e construa uma tabela de distribuição de frequências;
- b) Represente por meio de um gráfico a tabela obtida no item anterior;
- c) Calcule a média, a mediana e o desvio padrão para os níveis de chumbo (dica  $\sum x = 1126, \sum x^2 = 72302$ ).

2. Um jovem rapaz foi diagnosticado como tendo um determinado tipo de câncer. A proporção de jovens portadores deste tipo de câncer é de 1/10000. O teste ao qual o rapaz foi submetido tem 99% de probabilidade de produzir resultado positivo para um indivíduo que de

fato tem a doença, e 1% de probabilidade de produzir resultado positivo para um indivíduo que não tem o referido tipo de câncer. Qual a probabilidade do jovem a que se refere o enunciado ter câncer dado que o resultado do teste foi positivo?

3. Num experimento, doentes contaminados com cercaria, que é uma das formas do verme da esquistossomose, recebem um certo medicamento e observa-se a proporção de cura. De 200 pacientes medicados, 160 foram curados. Determine um intervalo de 90% de confiança para a proporção populacional de pacientes curados pelo medicamento.
4. Um pesquisador estudando pacientes com cirrose hepática queria investigar se o tratamento que utilizava conduzia a perda de peso. Para tanto, tomou uma amostra aleatória de 12 pacientes, registrando seus pesos (em kg) antes e depois da aplicação do tratamento. Formule hipóteses adequadamente e teste-as ao nível de 1% de significância.

Paciente	1	2	3	4	5	6	7	8	9	10	11	12
Antes	92,6	67,6	74,4	73	93,5	70,2	74,5	53,9	83	47,3	92,4	65,4
Depois	88	64,3	59,8	64	78,4	59,6	63,3	48,9	74	38	85,7	55,1

1. Que tipo de estudo você usaria para verificar a associação entre um fator de risco e um câncer raro? Justifique sua resposta.
2. O que é um grupo controle e qual sua importância nas pesquisas clínicas?
3. Um estudo foi conduzido comparando mulheres adolescentes que sofriam de bulimia com mulheres adolescentes com composição corporal e níveis de atividade física similares. Abaixo estão listadas as medidas de entrada calórica diária (  $\bar{X}$  ), registradas em kcal/kg, para as amostras de adolescentes de cada grupo:

Consumo calórico diário (kcal/kg)										
Bulímica	15,0	16,4	18,0	20,2	22,6	23,4	24,0	26,2	28,4	30,6
Saudável	20,8	22,4	26,4	28,2	30,0	31,0	33,4	35,6	38,0	40,2

- a) Obtenha as médias, as medianas e os desvios-padrões dos consumos calóricos das pacientes dos dois grupos. (Dica: para as pacientes bulímicas,  $\sum x = 225$ ,  $\sum x^2 = 5290$ ; para as pacientes saudáveis,  $\sum x = 306$ ,  $\sum x^2 = 9730$ ).
- b) Construa um boxplot para os consumos calóricos em cada um dos grupos. Represente-os numa única figura.
- c) Um valor típico do consumo calórico diário é maior para adolescentes que sofrem de bulimia ou adolescentes saudáveis? Qual grupo tem maior variabilidade nas medidas?
4. Um estudo registrou que a sensibilidade da mamografia como teste de triagem para detecção de câncer de mama é 0,85, enquanto sua especificidade é 0,80.
- a) Na população na qual a probabilidade de que uma mulher tenha câncer de mama é 0,0025, qual a probabilidade de que tenha câncer, se sua mamografia for positiva?
- b) Qual a probabilidade de um resultado falso positivo?
5. Doentes sofrendo de uma certa moléstia são submetidos a um tratamento intensivo cujo tempo de cura tem distribuição normal, com média 15 e desvio padrão 2 (em dias).
- a) Qual a probabilidade de um paciente submetido a esse tratamento apresentar tempo de cura inferior a 12 dias?
- b) Qual o tempo necessário para a recuperação de 25% dos pacientes (aqueles que se recuperam mais rapidamente)?

**Boa Prova!**

1. “Assim sendo, podemos concluir que os postos de saúde deste município, em 1995, contaram com a colaboração de 345 funcionários, distribuídos nas cinco regiões, a saber: região sul, 43; região norte, 102; região leste, 98; região oeste, 75; centro, 27. Na região sul, 38 eram do sexo feminino e na região norte, 87. No centro havia 20 funcionárias, na região leste trabalhavam 51 delas, enquanto que na região oeste, 35.”

- a. Com base nos dados fornecidos, complete a tabela de distribuição de frequências:

Região	Sexo		Total
	Feminino	Masculino	
Sul			
Norte			

Leste			
Oeste			
Centro			
Total			

- b. Construa um gráfico que permita comparar a distribuição dos funcionários nas cinco regiões do município quanto ao sexo.
2. Os dados apresentados na sequência referem-se à frequência cardíaca em batimentos por minuto de 49 pacientes submetidos a esforço físico. Os dados estão representados em ordem crescente no interior das colunas.

83	102	105	110	120	125	135
91	102	105	112	120	126	135
91	102	108	113	120	128	137
93	103	108	115	122	130	140
99	103	108	116	122	130	143
100	104	108	116	124	135	145
100	105	108	118	125	135	159

- a. Construa uma tabela de distribuição de frequências para os dados agrupados;
- b. Construa um gráfico de distribuição de frequências para com base na tabela obtida no item anterior;
- c. Calcule a média e o desvio para a frequência cardíaca dos pacientes, utilizando os dados não agrupados. (Dica:  $\sum x = 5671$ ;  $\sum x^2 = 668777$ ).
3. Um estudo na área de Medicina Ocupacional (MO) revelou que um em cada cinco motoristas de ônibus apresentava algum grau de lombalgia. Tomando ao acaso um grupo de 15 motoristas,
- Qual a probabilidade de dois apresentarem lombalgia?
  - Qual a probabilidade de três motoristas ou mais apresentarem lombalgia?
  - Qual o número esperado de motoristas com lombalgia? Qual a variância?
4. Suponha que o tempo de coagulação em seres humanos seja uma variável aleatória com distribuição normal de média igual a sete minutos e desvio padrão de dois minutos.
- Suponha que o tempo máximo aceitável para a coagulação do sangue de um indivíduo é de 10 minutos. Qual a probabilidade de uma pessoa escolhida ao acaso apresentar tempo de coagulação fora do limite aceitável?
  - Qual o valor que delimita 10% da população com menor tempo de coagulação?

1- Um hospital contrata novos médicos sempre que o tempo médio de espera de seus pacientes por um atendimento ultrapassa 20 minutos. Desejando investigar se há ou não necessidade de contratação de novos médicos, uma amostra de  $n$  pacientes terá o tempo de espera registrado, e a decisão de contratar ou não novos profissionais será tomada com base nessa amostra.

- Formule as hipóteses nula e alternativa para o problema apresentado;
- Como um erro do tipo I afetaria o hospital? E um erro do tipo II?
- Suponha que, com base nos resultados amostrais, tenha se obtido um nível descritivo (valor  $p$ ) igual a 0,03. Qual decisão deveria tomar o hospital, considerando um nível de 5% de significância?

2- O Centro de Acompanhamento Pré-Natal para dependentes de drogas Químicas da Escola de Medicina da Universidade de Northwestern – Chicago acompanhou a gravidez de 55 mulheres dependentes de cocaína. Desatas, apesar de todo o esforço do centro, apenas 19 conseguiram parar de usar a droga durante o 1º trimestre. A tabela abaixo apresenta os pesos dos recém-nascidos do grupo 1, filhos de mães que usaram cocaína apenas no primeiro trimestre de gravidez, e do grupo 2, filhos de mães que usaram cocaína durante toda a gravidez:

Informação	Grupo 1	Grupo 2
Tamanho da amostra	19	36
Média (em gramas)	3160	2829
Desvio padrão (em gramas)	453	512

- Forneça uma estimativa pontual para a diferença dos pesos médios de crianças filhas de mães que usaram cocaína durante toda a gravidez em relação às que usaram apenas no primeiro trimestre.
- Faça o teste estatístico adequado, ao nível de 5% de significância, para verificar a hipótese de que os filhos de mães que usaram cocaína durante toda a gravidez têm peso médio ao nascer diferente dos filhos de mães que usaram cocaína apenas no primeiro trimestre de gravidez;
- Estime a diferença dos pesos médios de crianças filhas de mães que usaram cocaína durante toda a gravidez em relação às que usaram apenas no primeiro trimestre por meio de um intervalo com 95% de confiança. Interprete o resultado.

3- Os seguintes dados são tamanhos (em dias) dos ciclos menstruais em uma amostra aleatória de 10 mulheres.

26, 24, 29, 33, 25, 26, 30, 31, 30, 28

- Teste se o número médio de dias do ciclo menstrual é igual ou não ao mês lunar (29,5 dias).
- Estime o número médio de dias do ciclo menstrual por meio de um intervalo com 95% de confiança. Interprete.
- Qual suposição acerca da distribuição do número de dias do ciclo menstrual você considerou válida para a resolução dos itens anteriores?

d) Caso você estivesse insatisfeito com a margem de erro do intervalo obtido no item b, e quisesse reduzi-la pela metade, dobrar o tamanho da amostra seria uma solução? Justifique.

1) Um laboratório afirma que dentre os doentes submetidos a um medicamento que comercializa, 2% apresentam determinado efeito colateral. Desconfiando que essa proporção é maior do que a apresentada pelo laboratório, um pesquisador decide avaliar uma amostra de doentes que fazem uso dessa medicação quanto à manifestação ou não desse efeito. Caso comprovada a desconfiança do pesquisador, o medicamento será imediatamente recolhido do mercado.

d) Formule as hipóteses nula e alternativa para o problema apresentado;

e) Qual a consequência de um erro do tipo I? E de um erro do tipo II?

f) Suponha que, com base nos resultados amostrais, tenha se obtido um nível descritivo (valor p) igual a 0,03. Considerando um nível de significância de 5%, há evidências suficientes para vetar a comercialização do medicamento?

a. Sabe-se que o tempo médio de cura para doentes tratados com um método convencional é conhecido, igual a 7 dias. Um método alternativo é proposto com a finalidade de diminuir o tempo de cura para essa doença. Em um experimento clínico, 8 pacientes com a doença receberam o tratamento alternativo, apresentando tempo médio de cura de 6 dias, com desvio padrão igual a 1 dia.

a) Formule e teste adequadamente as hipóteses de interesse, considerando um nível de 10% de significância;

b) Construa um intervalo de 90% de confiança para o tempo médio de cura referente ao tratamento alternativo.

3) A prescrição de anticoncepcionais orais deve ser efetuada após uma avaliação completa da paciente, uma vez que seu uso pode alterar diversas funções normais do organismo. Suponha que uma experiência com dois grupos de mulheres apresentou os seguintes resultados relativos à pressão arterial sistólica (PAS):

Grupo	Tamanho de amostra	Média (PAS, mmHg)	Desvio Padrão (PAS, mmHg)
Tratamento (T)	19	120,4	8,31
Controle (C)	25	115,6	16,22

Elabore um parecer conclusivo a respeito dos efeitos colaterais do anticoncepcional oral testado sobre a PAS, com base em um teste para a comparação de médias e na estimação intervalar da diferença entre as mesmas. Considere um nível de significância de 5%.

4) Os dados (adaptados) apresentados na sequência descrevem as frequências cardíacas de seis pacientes portadores de prótese valvar normofuncional antes e após a administração de 200mg de Amiodarona/dia:

Paciente	Antes	Depois
1	128	83
2	106	72
3	113	80
4	135	86
5	92	68
6	140	85

Com base nestes dados, testar se Amiodarona provoca redução na frequência cardíaca média, aos níveis de 5 e 1% de significância.

**Exercício 1** - Qual a diferença entre parâmetro, estimador e estimativa?

**Exercício 2** - Cite duas alternativas possíveis para se diminuir a margem de erro de uma estimativa intervalar.

**Exercício 3** - O que representa o nível de significância de um teste estatístico de hipóteses? E o nível descritivo?

**Exercício 4** - A fim de acelerar o tempo que um analgésico leva para penetrar na corrente sangüínea, um químico analista acrescentou certo componente à fórmula original, que acusava um tempo médio de 43 minutos. Em 36 observações com a nova fórmula, obteve-se um tempo médio de 42 minutos, com desvio padrão de 6 minutos.

- O que podemos concluir, ao nível de 5% de significância, sobre a eficiência do novo componente?
- Estime o tempo médio que a nova fórmula leva para penetrar na corrente sangüínea por meio de um intervalo com 95% de confiança.
- Caso você desejasse reduzir a margem de erro do intervalo apresentado no item anterior pela metade, em quantas vezes você deveria aumentar o tamanho da amostra?

**Exercício 5** - A tabela abaixo apresenta o tamanho da amostra, a média e a variância dos pesos de nascidos vivos de ambos os sexos.

Sexo	n	Média	Variância
Masculino	14	3,253	0,261
Feminino	14	3,130	0,265

- Estime pontualmente a diferença dos pesos médios para bebês dos dois sexos;
- Teste, a um nível de significância de 5%, a superioridade do peso médio de bebês do sexo masculino;



- c) Estime a diferença dos pesos médios de bebês dos dois sexos por meio de um intervalo com 95% de confiança;  
d) Qual suposição referente à distribuição dos pesos dos recém-nascidos você considerou válida para a resolução dos itens anteriores?

1. Num estudo sobre a associação entre tromboembolismo e tipo sanguíneo, participaram 220 usuárias de contraceptivo oral. Dessas mulheres, 55 tinham tromboembolismo. Quanto ao grupo sanguíneo, o tipo A foi o mais numeroso, com 83 mulheres, seguido dos grupos O e B, com 79 e 27 mulheres, respectivamente. Das pacientes saudáveis, 70 eram do grupo O, 51 do grupo A e 19 do grupo B.

- a) A partir dessas informações, preencha a tabela abaixo:

	TROMBOEMBOLISMO		
Grupo Sanguíneo	DOENTE	SADIA	TOTAL
A			
B			
AB			
O			
TOTAL			200

- b) Utilizando a tabela preenchida compare os dois grupos de mulheres (sadias e doentes) de forma gráfica.

2. Considere resultados de exames realizados em 48 cobaias (24 machos e 24 fêmeas):

Machos	35	35	35	37	37	38	38	39	39	40	40	40
	40	41	41	41	42	43	44	45	45	45	46	47
Fêmeas	38	38	42	42	44	46	48	49	50	51	51	51
	51	51	53	55	56	57	60	61	62	63	63	67

Complete a tabela de estatísticas descritivas para a variável em estudo, segundo o sexo da cobaia:

Grupo	Média	Desvio padrão	Coeficiente de variação	Mediana	Primeiro Quartil	Terceiro Quartil
Macho	40,54	3,54				
Fêmea	52,04	8,05				

3. O tempo (em meses) entre a remissão de uma doença e a recidiva de 48 pacientes de uma determinada clínica médica foi registrado. Os dados são apresentados abaixo:

2	2	3	4	4	4	4	7	7	7	8	9	9	10	12	15	15	15	16	18	18	22	22	24
2	2	3	3	4	4	5	5	6	6	7	7	7	7	8	8	8	8	10	10	11	11	12	18

Construa a distribuição de frequência. Calcule a média através dos dados originais e através dos dados tabelados.

4. Dentre as mulheres de 18 a 74 anos, nos EUA, a pressão sanguínea diastólica é normalmente distribuída com média 77 mm Hg e desvio padrão 11,6 mm Hg.

- Qual a probabilidade de que uma mulher selecionada aleatoriamente, desta faixa etária, tenha pressão sanguínea sistólica menor que 60 mm Hg?
- Qual a probabilidade de que ela tenha pressão maior do que 90 mm Hg?
- Qual a probabilidade de que tenha pressão entre 60 e 90 mm Hg?