

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE  
Organização Mundial da Saúde – Representação Brasil

**Módulo de Princípios de  
Epidemiologia para o Controle de  
Enfermidades (MOPECE)  
Medição das condições de saúde  
e doença na população**

3

$$RR = \frac{I_E \cdot \frac{a}{a+b}}{I_E \cdot \frac{c}{c+d}}$$

$$OR = \frac{(a+0.5) \times (d+0.5)}{(b+0.5) \times (c+0.5)}$$



Organização Pan-Americana da Saúde

# Módulo de Princípios de Epidemiologia para o Controle de Enfermidades (MOPECE)

Módulo 3: Medição das condições de saúde  
e doença na população

Brasília – DF  
2010

© 2010 Organização Pan-Americana da Saúde.

Todos os direitos reservados. É permitida a reprodução parcial ou total dessa obra, desde que citada a fonte e que não seja para venda ou qualquer fim comercial.

Versão preliminar: traduzida para o português e adaptada, 2010

*Módulos de Principios de Epidemiología para el Control de Enfermedades* foi elaborado pelo Programa Especial de Análises de Saúde do Escritório Central da Organização Pan-Americana da Saúde (Washington, DC-EUA) em 2001. ISBN: 92 75 32407 7.

A versão em português, que corresponde aos Módulos de Principios de Epidemiologia para o Controle de Enfermidades, foi revisada pela Unidade Técnica de Informação em Saúde, Gestão do Conhecimento e Comunicação da OPAS/OMS no Brasil e pelo Ministério da Saúde por meio do Departamento de Análise de Saúde e pela Coordenação-Geral de Desenvolvimento da Epidemiologia em Serviços.

*Elaboração, distribuição e informações:*

ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE – REPRESENTAÇÃO BRASIL

Setor de Embaixadas Norte, Lote 19

CEP: 70800-400 Brasília/DF – Brasil

<http://www.paho.org/bra>

MINISTÉRIO DA SAÚDE

Secretaria de Vigilância em Saúde (SVS)

Esplanada dos Ministérios, Bloco G

CEP: 70058-900 Brasília/DF – Brasil

<http://www.saude.gov.br>

*Revisão técnica:*

José Moya, Oscar J. Mujica e Giselle Moraes Hentzy (OPAS/OMS)

Maria Regina Fernandes, Marta Helena Dantas e Adauto Martins Soares Filho (SVS/MS)

*Colaboração:*

Jarbas Barbosa, Fátima Marinho, Oscar J. Mujica, José Escamilla, João Baptista Risi Junior,

Roberto Becker (OPAS/OMS)

*Capa, Projeto Gráfico e Diagramação:*

All Type Assessoria Editorial Ltda

Impresso no Brasil / *Printed in Brazil*

#### Ficha Catalográfica

---

Organização Pan-Americana da Saúde

Módulos de Principios de Epidemiología para o Controle de Enfermidades. Módulo 3: medida das condições de saúde e doença na população / Organização Pan-Americana da Saúde. Brasília : Organização Pan-Americana da Saúde ; Ministério da Saúde, 2010.

94 p.: il. 7 volumes.

ISBN 978-85-7967-021-3

Título original: *Módulos de Principios de Epidemiología para el Control de Enfermedades*.

1. Saúde Pública – Epidemiologia. 2. Educação Profissional em Saúde Pública. I. Organização Pan-Americana da Saúde. II. Ministério da Saúde. III. Título.

---

NLM: WC 503.4

Unidade Técnica de Informação em Saúde, Gestão do Conhecimento  
e Comunicação da OPAS/OMS no Brasil

## Sumário

---

<b>Apresentação</b> .....	<b>5</b>
<hr/>	
<b>Conteúdos e objetivos</b> .....	<b>7</b>
<hr/>	
<b>Medição de saúde e doença na população.</b> .....	<b>8</b>
<hr/>	
<b>Tipos de dados e sua tabulação</b> .....	<b>12</b>
<hr/>	
<b>Apresentação gráfica de dados</b> .....	<b>17</b>
<hr/>	
<b>Medidas de resumo de uma distribuição</b> .....	<b>22</b>
<hr/>	
<b>Medidas de tendência central.</b> .....	<b>23</b>
<hr/>	
<b>Medidas de dispersão</b> .....	<b>29</b>
<hr/>	
<b>Medidas de frequência</b> .....	<b>35</b>
Prevalência e incidência .....	36
<hr/>	
<b>Comparação da frequência da doença e medidas de associação</b> .....	<b>62</b>
<hr/>	
<b>Medidas de associação</b> .....	<b>63</b>
Comparação de duas Proporções: O Teste do Chi-quadrado .....	65
Medidas da força de associação .....	68
Risco Relativo .....	68
Razão de Chances ( <i>Odds Ratio</i> ) .....	75

---

**Apêndice estatístico - epidemiológico . . . . .79**

Comparação das Médias: O Teste Z . . . . .79  
Intervalos de Confiança para médias e proporções . . . . .80  
Medidas de impacto potencial . . . . .84  
Riscos atribuíveis . . . . .84  
Frações atribuíveis . . . . .85  
Padronização de taxas . . . . .89

---

**Referências bibliográficas . . . . .93**

## Apresentação

A Organização Pan-Americana da Saúde no Brasil (OPAS) tem grande satisfação em apresentar os Módulos de Princípios de Epidemiologia para o Controle de Enfermidades (MOPECE) na versão traduzida para a língua portuguesa.

O MOPECE é um instrumento de capacitação em epidemiologia básica, voltado para profissionais de saúde, especialmente aqueles que atuam nos serviços de saúde locais, que tem por finalidade promover o conhecimento e a aplicação prática dos conteúdos epidemiológicos no enfrentamento dos problemas de saúde local, assim como no apoio ao planejamento e gestão em saúde.

A primeira edição do MOPECE, lançada na década 80, foi escrita em espanhol e teve ampla divulgação na região das Américas. Em 2001, mediante a incorporação de novos conceitos e avanços no campo da epidemiologia, foi proposta uma segunda edição.

Para essa publicação, além da tradução da segunda edição para a língua portuguesa, foram incluídas informações de relevância para a saúde pública, tais como: Orientações sobre o novo Regulamento Sanitário Internacional (RSI-2005), descrição de uma investigação de surto de toxoplasmose realizada por profissionais brasileiros, como parte do Programa de Treinamento em Epidemiologia Aplicado aos Serviços do Sistema Único de Saúde (EPISUS), entre outras.

Este trabalho é resultado da cooperação técnica entre a OPAS/OMS e a Secretaria de Vigilância em Saúde (SVS) do Ministério da Saúde do Brasil com o objetivo de favorecer o aperfeiçoamento dos profissionais que compõe a força de trabalho do Sistema Único de Saúde (SUS), especialmente aqueles que atuam no programa de saúde da família e em centros de formação em saúde. Em adição, essa publicação contribui com a estratégia de cooperação internacional (sul-sul); particularmente com os países de língua portuguesa.

*Dr. Jarbas Barbosa*

Gerente da Área de Vigilância da Saúde e  
Prevenção e Controle de Doenças da OPAS

*Eng. Diego Victoria*

Representante da OPAS/OMS no Brasil





## Conteúdos e objetivos

Esse Módulo apresenta as medidas comuns de quantificação das condições de saúde e doença, suas aplicações e limitações; a ordenação e apresentação tabular e gráfica de dados epidemiológicos; as medidas de resumo e de associação. Esse Módulo está focado nos métodos para a apresentação e análise de dados quantitativos.

Os objetivos do presente Módulo são:

- Descrever as formas de medida da saúde e suas aplicações.
- Identificar as formas e usos da apresentação tabular e gráfica de dados.
- Descrever, calcular, interpretar e aplicar as medidas de resumo de dados.
- Descrever, calcular, interpretar e aplicar as medidas de associação.

## Medição de saúde e doença na população

Existem diversas formas de medir a saúde, dependendo de qual é a sua definição; uma definição ampla mediria o nível de saúde e bem-estar, a capacidade funcional, a presença e causas de doenças e óbito e a expectativa de vida das populações (Donaldson, 1989). Existem diferentes medidas e indicadores de bem-estar (social e econômico) na saúde e foram desenvolvidos certos índices de “saúde positiva” (Alleyne, 1998), tanto com fins operacionais, como para a investigação e promoção de condições saudáveis, em dimensões tais como a saúde mental, auto-estima, satisfação com o trabalho, exercício físico, etc. A coleta de dados e a estimativa de indicadores têm como finalidade gerar, de forma sistemática, evidências que permitam identificar padrões e tendências que ajudem a empreender ações de proteção e promoção da saúde e de prevenção e controle de doenças na população.

Entre as formas mais úteis e comuns de medir as condições gerais de saúde da população, destacam-se os censos nacionais, que são feitos a cada década em vários países. Os censos proporcionam a contagem periódica da população e a descrição de várias das suas características, cuja análise permite fazer estimativas e projeções.

Para permitir as comparações ao longo do tempo numa mesma população ou entre populações diferentes, são necessários procedimentos de medição padronizados.

**Medição:** é o procedimento de aplicação de um padrão a uma variável ou a um conjunto de valores.

A medição do estado de saúde requer sistemas harmonizadores e unificados como a Classificação Estatística Internacional de Doenças e Problemas Relacionados à Saúde (CID), na sua Décima Revisão, cujos XXII capítulos iniciam com certas doenças infecciosas e parasitárias (A00-B99) e terminam com o capítulo referente aos códigos para propósitos especiais (U00-U99).

Os indicadores de saúde medem na população diferentes aspectos relacionados com a função ou incapacidade, a ocorrência de doença ou óbito, bem como os aspectos relacionados com os recursos e desempenho dos serviços de saúde.

Os indicadores de saúde funcional tratam de medir o impacto dos problemas de saúde na vida diária, por exemplo, a capacidade para realizar atividades cotidianas, lesões e acidentes domésticos e no local do trabalho, e anos de vida livres de incapacidade. Os dados são obtidos geralmente através de inquéritos e registros de incapacidade. Os índices de qualidade de vida incluem variáveis de função como a atividade física, a presença de dor, o nível do sono, de energia e o isolamento social.

Os indicadores de **morbidade** medem a frequência de problemas de saúde específicos como infecções, cânceres, acidentes de trabalho, etc. As fontes de dados costumam ser registros de hospitais e serviços de saúde, notificação de doenças sob vigilância e inquéritos de soroprevalência e de auto-relato de doenças, entre outros. É necessário mencionar que as doenças crônicas, pela sua longa duração, requerem o monitoramento das etapas clínicas, pelo que é preferível contar com os registros de doença (ex: câncer, defeito congênito etc) (Newcomer, 1992).

Os indicadores de **mortalidade** geral ou por causas específicas permitem comparar o nível geral de saúde e identificar causas de mortalidade relevantes como acidentes, tabagismo, etc. O registro da mortalidade requer o atestado de óbito, para o qual é usado a Declaração de Óbito. A mortalidade se apresenta geralmente como números absolutos, proporções, ou taxas por idade, sexo e causas específicas.

Além da medição do estado de saúde, também é necessário medir o **desempenho dos serviços de saúde**. Tradicionalmente, essa medição é focada para os insumos e serviços; atualmente, considera-se preferível medir os processos e funções dos serviços de saúde (Turnock, 1997).

Conjuntamente com os indicadores mencionados, a mensuração na saúde requer a disponibilidade de dados sobre características relevantes da **população** (variáveis), tais como seu tamanho, composição, estilos de vida, classes sociais, eventos de doenças, nascimentos e óbitos.

Os dados para a medição da saúde provêm de diversas fontes, motivo pelo qual devem ser considerados os aspectos relacionados com a invalidez, qualidade, integridade e cobertura dos próprios dados e suas fontes. Os dados, quantitativos ou qualitativos, que se obtêm e se registram dos serviços de saúde e das estatísticas vitais representam a “matéria prima” para o trabalho epidemiológico. Quando os dados são incompletos ou inconsistentes, serão obtidas medidas enviesadas ou inexatas, sem importar a sofisticação da análise epidemiológica, e as intervenções derivadas do seu uso não serão efetivas. A deficiente cobertura dos serviços, em amplos setores da população de vários países, limita a geração de informação útil e necessária para resolver os problemas de saúde que atingem de forma específica as suas comunidades. Ainda quando os dados estiverem disponíveis e sejam confiáveis, sua utilização para a gestão em saúde pode ser insuficiente.

Com o propósito de responder às necessidades de contar com um conjunto de dados válidos, padronizados e consistentes dos países das Américas, a Organização Pan-Americana da Saúde (OPAS) trabalha desde 1995 na Iniciativa Regional de Dados Básicos de Saúde. Está incluída nessa fonte uma série histórica de 117 indicadores demográficos, socioeconômicos, de morbidade, mortalidade, de recursos, acesso e cobertura de serviços de saúde, dos 48 países e territórios da região.

Uma vez contados os dados e calculados os indicadores de saúde, uma das dificuldades apresentada nos serviços de saúde está relacionada com as limitações do manuseio correto da informação numérica, sua análise e interpretação, funções que requerem o uso dos princípios da epidemiologia e da bioestatística. No entanto, no âmbito em que ocorrem os problemas e onde eles são solucionados, os procedimentos e técnicas para obtenção, medição, processamento, análise, interpretação dos dados e uso das informações ainda não estão plenamente desenvolvidos.

Para a correta tomada de decisões em todos os níveis dos serviços de saúde, baseada na informação pertinente, é necessária a capacitação permanente da equipe local de saúde e das suas redes na coleta, manuseio, análise e interpretação de dados epidemiológicos.

A quantificação dos problemas de saúde na população requer procedimentos e técnicas estatísticas diversas, algumas delas de relativa complexidade. Dadas as características de múltiplos fatores dos problemas de saúde, as técnicas qualitativas são também valiosas para aproximar-se do conhecimento dos determinantes da saúde. É por isso que existe a necessidade de incorporar, de forma dialética, métodos e técnicas quantitativas e qualitativas que permitam estudar os diversos componentes dos objetos de estudo.

Na análise quantitativa o uso de programas informatizados facilita o manuseio e a análise de dados, mas não se deve superestimar seu alcance e aplicações. Sua utilidade é maior quando são estabelecidas redes de colaboração e sistemas de informação em saúde, que permitem o manuseio eficiente de grandes bases de dados e geram informação oportuna e útil para a tomada de decisões. Um programa informatizado reduz notavelmente o tempo de cálculo, processamento e análise dos dados, mas é o trabalho humano o que aporta resultados racionais e válidos para o desenvolvimento dos objetivos de saúde pública. Existem dois pacotes de programas de cálculo desenhados especificamente para a saúde que facilitam o armazenamento, processamento e análise de informação epidemiológica: Epi-Info, produzido pelo Centro de Prevenção e Controle de Doenças dos Estados Unidos (CDC) e a Organização Mundial da Saúde (OMS), e o Epi-Dat, desenvolvido pela OPAS e a Xunta de Galicia, Espanha. Longe de competir entre si, os pacotes de programas, de grande uso e de livre distribuição, oferecem processos e rotinas de manuseio e análise epidemiológica de dados que são complementares.

Em um sentido amplo, podemos considerar que o trabalho da saúde pública parte da constatação de uma realidade de saúde não desejável em uma população e aponta para conseguir mudanças sociais, deliberadas e sustentáveis nessa população. Nesse sentido, e do ponto de vista metodológico, a epidemiologia como toda ciência tem exigência de método, desde uma perspectiva estatística. O **foco epidemiológico** consiste basicamente em: i) a *observação* dos fenômenos de saúde e doença na população; ii) a *quantificação* dos mesmos em frequências e distribuições; iii) a *análise* das frequências e distribuições de saúde e de seus determinantes; e iv) a *definição de cursos de ação* apropriados. Esse

processo de observar-quantificar-comparar-propor serve também para avaliar a efetividade e o impacto das intervenções em saúde, para construir novos modelos que descrevam e expliquem as observações e para utilizá-los na predição de novos fenômenos.

Em resumo, em todo esse processo, os procedimentos e técnicas de quantificação são de grande relevância, e a capacitação da equipe local de saúde nestes aspectos do enfoque epidemiológico é, conseqüentemente, fundamental.

## Tipos de dados e sua tabulação

A quantificação do estado de saúde e padrões de doença na população requer métodos e técnicas que permitam coletar dados de forma objetiva e eficiente; converter os dados em informação para facilitar sua comparação e simplificar sua interpretação; e transformar a informação em conhecimento relevante para as ações de controle e prevenção. Para conhecer os grupos de população que apresentam maior número de casos, os lugares com maior incidência ou prevalência de determinadas doenças e o momento em que ocorrem os eventos na saúde, devem-se aplicar sistematicamente alguns processos básicos de manuseio de dados ou variáveis.

Um dos primeiros passos no processo de medição do estado de saúde na população é a definição das variáveis que o representam ou caracterizam.

**Variável:** qualquer característica ou atributo que pode assumir valores diferentes.

As variáveis podem ser de dois tipos, qualitativa e quantitativa. Denominamos variáveis qualitativas aquelas que são atributos ou propriedades. As variáveis quantitativas são aquelas nas quais o atributo mede-se numericamente e, por sua vez, podem ser classificadas em discretas e contínuas. As **variáveis discretas** assumem valores que são sempre números inteiros; por exemplo, o número de filhos de um casal, o número de dentes com cárie, o número de leitos de um hospital, o número de hemácias por campo, o pulso, etc, que somente podem tomar valores de um conjunto finito. As **variáveis contínuas** podem tomar tantos valores quanto permita a precisão do instrumento de medição; por exemplo, o peso ao nascer de um bebê de 2.500 gramas pode ser medido com maior precisão, como 2.496,75 gramas, se a nossa balança assim o permitir.

As variáveis também podem ser classificadas conforme o nível ou tipo de medição que possamos aplicar. Assim, podem-se diferenciar quatro níveis de medição das variáveis: nominal, ordinal, de intervalo e proporcional ou de razão. Uma **variável nominal** tem categorias às quais são atribuídas nomes que não têm nenhuma ordem entre eles; por exemplo, o sexo. A categoria “homem” não tem relação de ordem sobre a categoria “mulher”. As variáveis nominais não têm de ser necessariamente dicotômicas (duas categorias) podem ter várias categorias, como, por exemplo, o estado civil (solteiro, casado, divorciado, viúvo, união livre) ou o grupo sanguíneo segundo o sistema ABO (A, B, AB e O). O fato de mudar a ordem não tem nenhuma implicação na análise dos dados.

Agora suponhamos que nos seja feita uma pergunta sobre a qualidade de um curso que acabamos de realizar e nos oferecem as seguintes opções de resposta: *muito ruim, ruim, regular, bom e excelente*. Essa classificação tem uma ordem: *excelente* é melhor que *bom*, *bom* que *regular* e assim sucessivamente; no entanto, a “distância” que existe entre

*excelente e bom* não tem que ser a mesma que entre *ruim e muito ruim*. Estamos ante uma **variável ordinal**, cujas categorias têm uma ordem, ainda que as diferenças entre elas possam não ser iguais. Outros exemplos de variáveis ordinais são os estágios de um câncer (I, II, III e IV) ou os resultados de uma cultura de laboratório (–, +, ++, +++).

Outro modo de medição de variáveis é o de intervalo. Uma **variável de intervalo** tem distâncias iguais entre seus valores e uma característica fundamental: *o zero é arbitrário*. O exemplo típico de variável de intervalo é a temperatura corporal. Existe a mesma diferença entre 37°C e 39°C e 38°C e 40°C (ou seja, 2°C). No entanto, não podemos dizer que uma temperatura de 60°C seja “três vezes mais quente” que uma de 20°C. Tampouco podemos concluir que um indivíduo com um coeficiente de inteligência de 120 é o dobro de inteligente que outro com coeficiente 60.

Por último, se a variável tivesse um ponto de origem que é o valor *zero absoluto*, estaríamos falando de uma **variável proporcional ou de razão**. Essa tem intervalos iguais entre valores e o ponto de origem zero. O peso e a altura são exemplos típicos de variáveis proporcionais, 80 kg é o dobro de 40 kg, e há a mesma diferença entre 50 e 35 kg, que entre 105 e 90 kg. Nesse nível, pode-se somar, diminuir, multiplicar e dividir.

No seguinte esquema, sintetiza-se a relação entre os diferentes níveis de medição das variáveis.

Tipo de variável	Valores
Nominal	categorias com nome convencional
Ordinal	como as nominais, + <i>categorias ordenadas</i>
De Intervalo	como as ordinais, + <i>intervalos iguais</i>
Proporcional ou de Razão	como as de intervalo, + <i>zero significativo</i>

Adaptado de Norman y Streiner, 1996.

Os dados sobre casos de doenças atendidas ou notificadas pelo centro ou serviço de saúde podem vir de uma lista de nomes, idades, sexo, etc, da qual se pode obter o número total ou frequência de casos. Se for desejável agrupar os casos conforme algumas características dos mesmos, a tarefa de identificar os grupos populacionais com maiores problemas fica simplificada. Por exemplo, entre as características da **pessoa**, podem-se agrupar os casos conforme sua idade, sexo ou etnia. A partir dessa lista de dados, pode-se, também, identificar os casos que ocorreram em determinado **lugar** e comparar a frequência da doença em diferentes lugares ou áreas geográficas. Por último, pode-se examinar o número de casos conforme o momento em que ocorreram ou foram notificados. Quanto à distribuição dos casos no **tempo**, pode-se agrupar em dias, semanas, meses ou qualquer outro período de tempo que se considere adequado. Isso permite saber em que momento se apresentou o maior número de casos, quando começou a au-

mentar e quando começou a diminuir. É necessário lembrar que a frequência é o número de vezes que se repete um valor da mesma variável.

Os dados agrupados conforme determinadas características (idade, sexo, residência, classe social, etc) podem ser apresentados em tabelas e/ou gráficos; isso facilita os cálculos e a comparação e interpretação dos resultados como é apresentado na seguinte Tabela.

**Tabela 3.1** Distribuição dos casos de gastroenterites conforme a idade. Centro de Saúde A, maio de 2000.

Idade (anos)	Casos (Nº)	Porcentagem %
< 1	63	36,4
1	55	31,8
2	25	14,4
3	14	8,1
4	5	2,9
5 e mais	11	6,4
Total	173	100,0

### Exemplo para a elaboração de uma tabela:

A seguinte lista corresponde às idades em anos de 120 pessoas atingidas pela malária durante o verão de 2001, em uma ilha do Caribe.

27	32	58	44	24	32	29	50
28	36	38	48	38	47	29	39
40	37	35	36	36	36	42	45
32	36	48	42	46	35	32	54
30	29	36	44	30	36	27	37
35	33	38	63	37	53	35	46
36	27	34	32	46	38	43	20
25	41	27	53	40	31	47	43
29	49	50	34	47	36	38	24
30	51	43	46	38	49	47	30
29	33	54	40	28	63	36	41
46	51	49	37	41	37	39	38
30	35	36	34	43	43	37	55
29	44	38	42	43	35	42	50
35	47	32	54	41	41	35	40

A maior idade é de 63 anos e a menor é de 20 anos. Se agruparmos estas idades em classes, por intervalos de 4 anos, teremos 11 grupos de idade, aos quais atribuiremos as suas respectivas frequências, como é mostrado a seguir:



20 – 23 anos	(um caso)
24 – 27 anos	(sete casos)
28 – 31 anos	(14)
32 – 35 anos	(19)
36 – 39 anos	(27)
40 – 43 anos	(19)
44 – 47 anos	(14)
48 – 51 anos	(10)
52 – 55 anos	(6)
56 – 59 anos	(1)
60 – 63 anos	(2)

Finalmente, preparamos a apresentação tabular dessa informação que se denomina tabela resumo de frequências, útil porque apresenta a distribuição da variável que estamos observando.

**Distribuição:** é o resumo completo das frequências dos valores ou categorias da medição realizada. A distribuição mostra quantos ou que proporção do grupo se encontra em um determinado valor ou categoria de valores dentro de todos os possíveis que a medida quantitativa possa ter.

Em nosso exemplo, a distribuição do número de casos da doença segundo grupos de idade se apresenta na tabela 3.2. A distribuição do número de casos (frequência absoluta) acompanha a porcentagem de cada grupo (frequência relativa simples) e a porcentagem acumulada (frequência relativa acumulada), que costumam aportar informação adicional útil. Por exemplo, 22,5% dos casos correspondeu a pessoas de 36 a 39 anos de idade e mais da metade dos casos (56,6%) tinham menos de 40 anos.

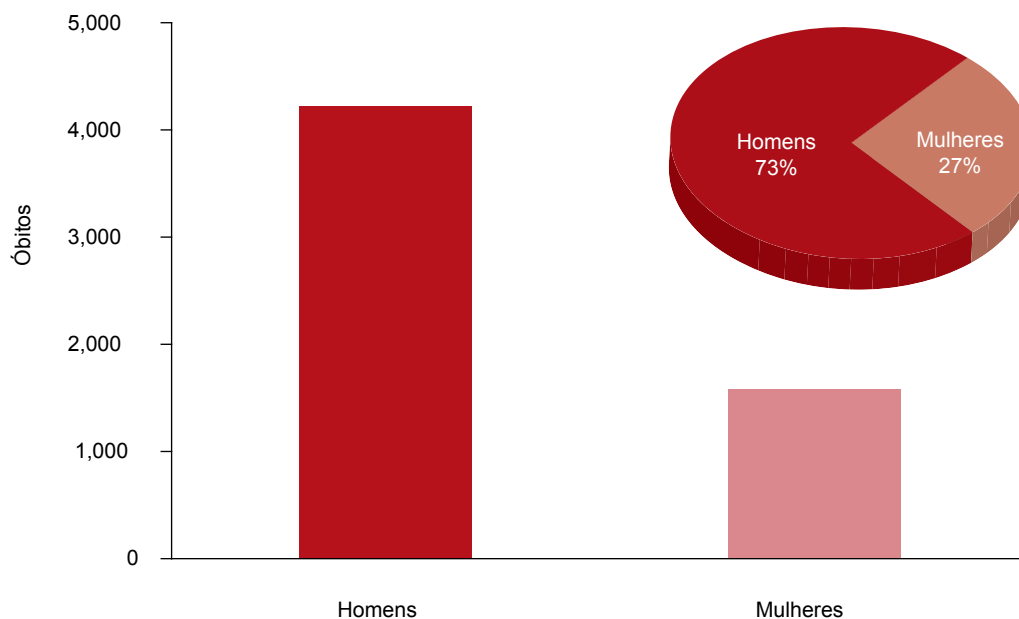
**Tabela 3.2** Distribuição de casos de malária por grupos de idade. Ilha do Caribe, verão de 2001

Faixa etária (em anos)	Número de casos (frequência absoluta)	Porcentagem (%)	Porcentagem acumulada (%)
20 - 23	1	0,8	0,8
24 - 27	7	5,8	6,6
28 - 31	14	11,7	18,3
32 - 35	19	15,8	34,1
36 - 39	27	22,5	56,6
40 - 43	19	15,8	72,4
44 - 47	14	11,7	84,1
48 - 51	10	8,4	92,5
52 - 55	6	5,0	97,5
56 - 59	1	0,8	98,3
60 - 63	2	1,7	100,0
Total	120	100,0	

## Apresentação gráfica de dados

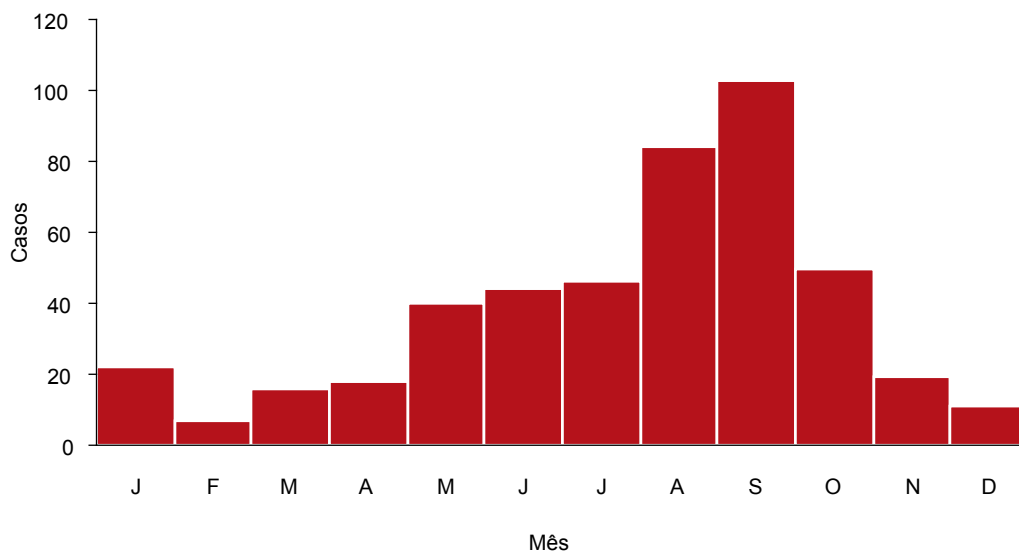
A distribuição de variáveis qualitativas, bem como as quantitativas discretas, costuma ser apresentada graficamente através de **diagramas de barras** ou por **gráficos de setores**, seja como frequências absolutas ou relativas, como é demonstrado a seguir:

**Gráfico 3.1** Distribuição de óbitos por suicídio conforme sexo. Lugar X, 1995-2000



As variáveis quantitativas contínuas são representadas graficamente por meio de **histogramas** e **polígonos de frequência**. Ainda que sejam parecidos aos diagramas de barras, nos histogramas, as barras se dispõem de forma adjacente, precisamente para ilustrar a continuidade e distribuição da variável representada. No eixo “x”, se localiza a variável contínua e no eixo “y” se representa a frequência.

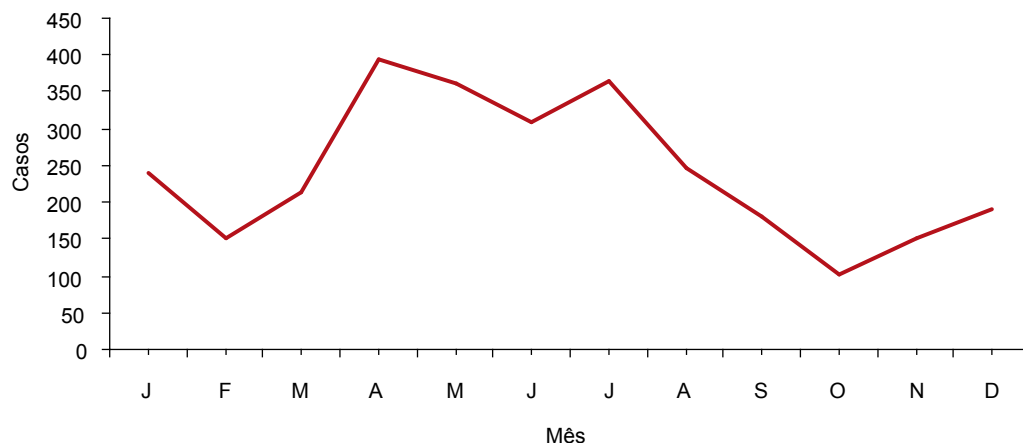
**Gráfico 3.2.** Casos notificados de Giardíase por mês de ocorrência, Lugar X, 2000 (total de casos notificados = 460)



No Gráfico 3.2, as categorias (meses) da variável tempo, uma variável contínua, estão representadas no eixo horizontal e o número de casos de giardíase notificados estão no eixo vertical. A altura de cada barra representa a frequência absoluta de casos (pode também ser a frequência relativa) em cada uma das categorias da variável tempo, chamadas também de intervalos de classe. Note que os intervalos de classe devem ser de igual tamanho.

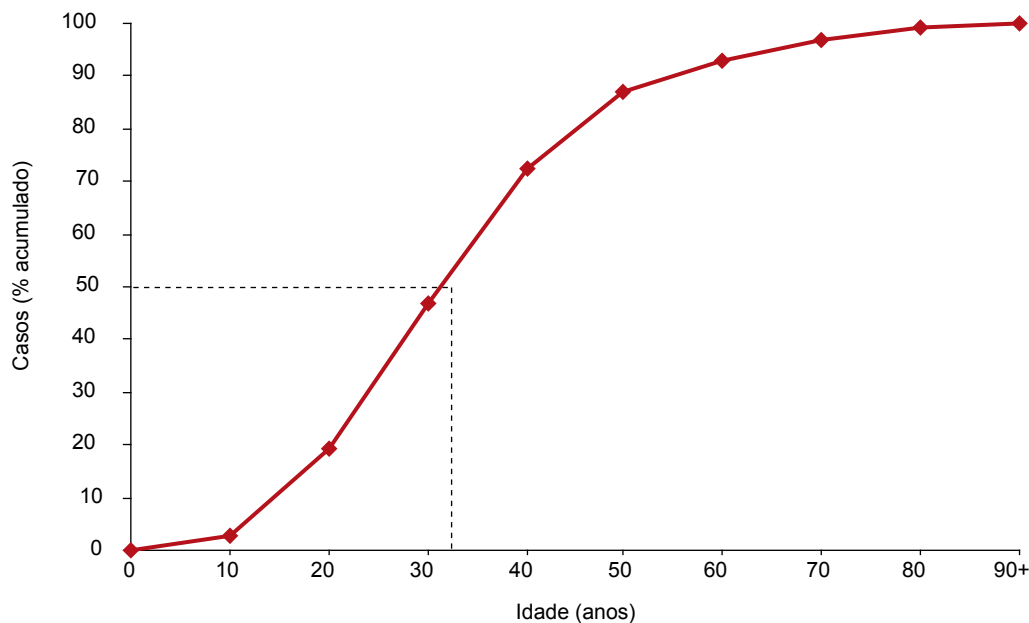
O polígono de frequências também permite representar graficamente a distribuição de uma variável e ser construído unindo com linhas retas os pontos médios do extremo superior de cada barra de um histograma (Gráfico 3.3). É particularmente útil para visualizar a forma e simetria de uma distribuição de dados e para apresentar simultaneamente duas ou mais distribuições.

**Gráfico 3.3** Casos de dengue por mês de início dos sintomas. Lugar Y, 2000



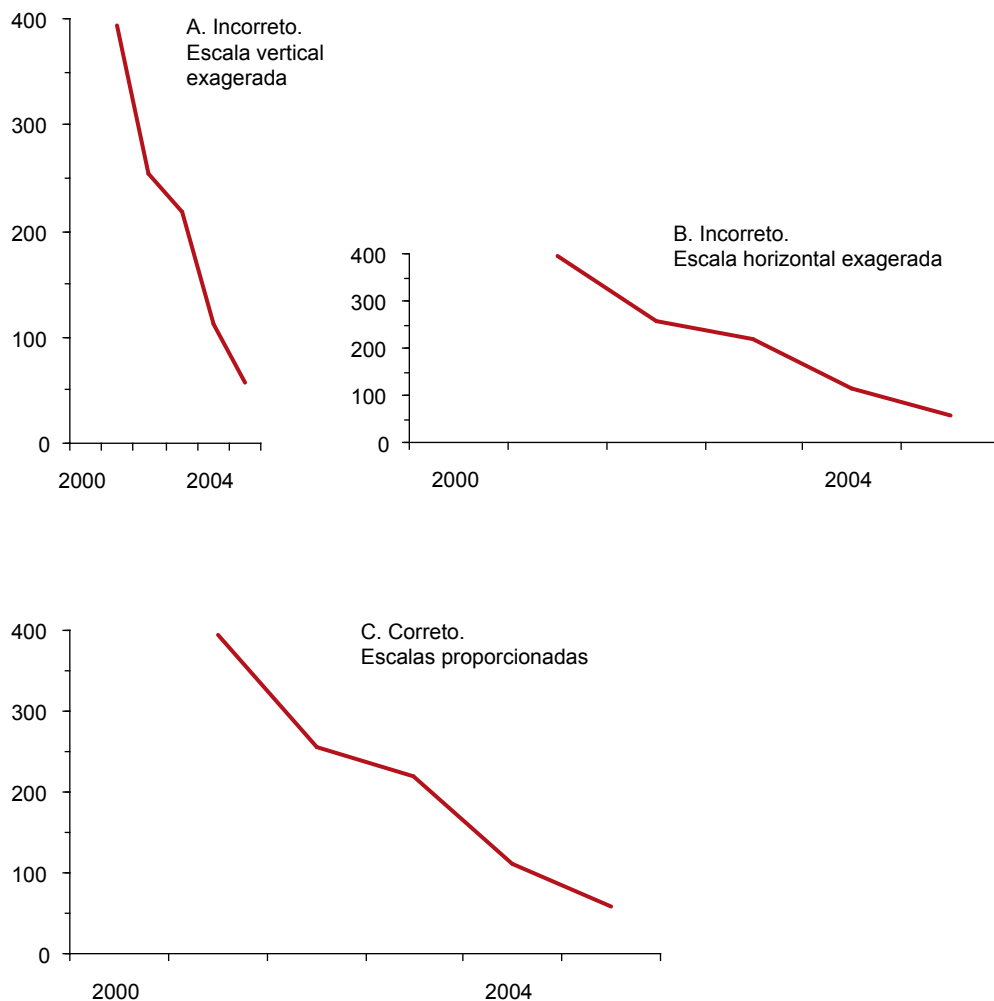
Uma variante do polígono de frequências é chamada “**ogiva percentual**”, que é um polígono de frequências relativas acumuladas. Cada ponto desse polígono representa a porcentagem acumulada de casos em cada intervalo de classe e, portanto, vai de zero a 100%. A ogiva percentual permite identificar de forma gráfica o valor correspondente à mediana (isto é, 50%) da distribuição, por exemplo, 32 anos no Gráfico 3.4.

**Gráfico 3.4** Casos de tuberculose por idade. Lugar Z, 2000



É importante manter a proporção das escalas do gráfico, já que de outra maneira esse pode dar impressões errôneas (Gráfico 3.5, A, B, C).

**Gráfico 3.5** Efeito das dimensões das escalas nos gráficos



Apesar de não existir uma regra explícita sobre a proporcionalidade entre as escalas de um gráfico, recomenda-se que a razão entre a escala horizontal em relação à vertical se aproxime a 1,6: 1 (a clássica “razão de ouro”). Algumas vezes, a posição 1,2 a 2,2 aparece como referência apropriada para a razão entre o eixo horizontal em relação ao eixo vertical.

O seguinte esquema resume os tipos básicos de gráficos mais apropriados segundo cada tipo de variável.

Tipo de variável	Tipo de gráfico
Nominal	diagrama de barras gráfico de setores
Ordinal	diagrama de barras (*) gráfico de setores
De Intervalo	diagrama de barras (*) histograma gráfico de setores (**) polígonos de frequências (simples e acumuladas)
Proporcional ou de Razão	diagrama de barras (*) histograma gráfico de setores (**) polígonos de frequências (simples e acumuladas)

(\*) mantendo a ordem das categorias

(\*\*) em variáveis discretas ou categorizadas

Uma forma de apresentação gráfica do comportamento histórico de uma doença e os limites de variabilidade esperados sobre a sua ocorrência no futuro, denomina-se **corredor ou canal endêmico**. Essa ferramenta da vigilância (Módulo 4) utiliza polígonos de frequência e permite visualizar o comportamento secular (isto é, no tempo) das doenças em um determinado território e orientar a decisão sobre a necessidade de desencadear ações de prevenção e controle.

## Medidas de resumo de uma distribuição

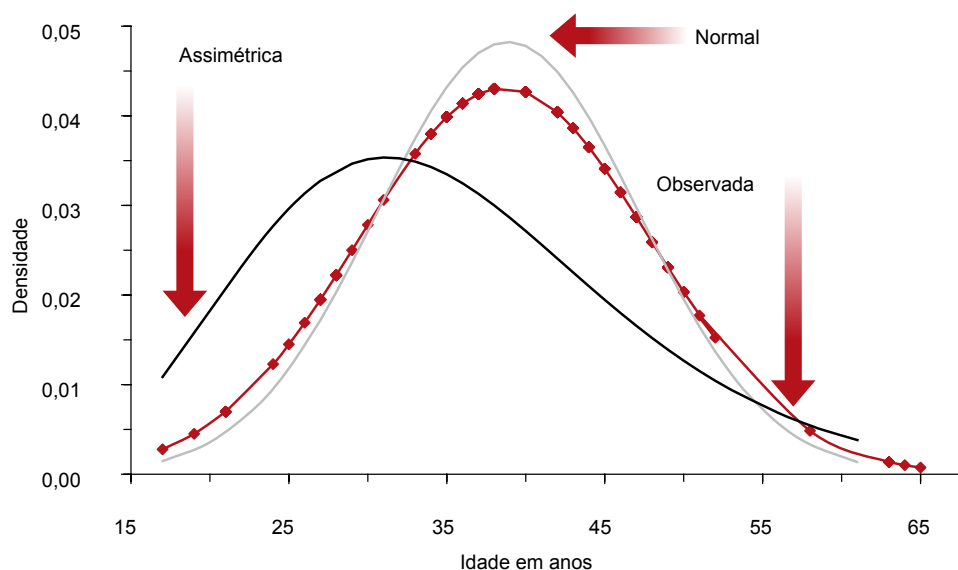
Muitas vezes é necessário utilizar um valor resumo que represente a série de valores em seu conjunto, isso é, sua distribuição. Para os dados de variáveis qualitativas e as quantitativas discretas, a proporção ou porcentagem, a razão e as taxas são algumas medidas típicas de resumo. Para os dados de variáveis quantitativas contínuas, no entanto, há medidas que resumem sua tendência para um valor médio (medidas de tendência central) e outras que resumem seu grau de variabilidade (medidas de dispersão). Cada uma proporciona informação complementar e útil para a análise epidemiológica.



## Medidas de tendência central

As medidas de tendência central dos dados são a moda, a mediana e a média. A seleção das medidas depende do tipo de dados e propósitos. Os valores de muitas variáveis biológicas, como altura, se distribuem de modo simétrico. Outras variáveis, como a mortalidade e a taxa de ataque de uma epidemia, têm distribuição assimétrica. Para distribuições normais (exemplo de distribuição simétrica), a média, a mediana e a moda são idênticas. Para distribuições assimétricas, a mediana representa melhor o conjunto de dados, apesar de que a média tem melhores propriedades para a análise estatística e testes de significância. O Gráfico 3.6 ilustra estas distribuições, usando a série de idades em anos dos 120 casos de malária, 2001, ilha do Caribe, apresentada anteriormente.

**Gráfico 3.6** Distribuição normal, observada e assimétrica da idade



Para indicar que, por exemplo, a rubéola tem um período de incubação de 15 a 21 dias (geralmente 18 dias), foi necessário observar uma grande quantidade de casos, desde o primeiro contato de uma pessoa sadia com um doente até o surgimento dos sinais e sintomas e medir esse período em dias. A diferença entre o limite inferior (15 dias) e o superior (21 dias) é conhecido como amplitude e se considera junto com o desvio-padrão e a variância, uma medida de dispersão de dados, como será visto mais adiante. A tabela 3.3 apresenta o período de incubação da rubéola, em dias, observado em 11 crianças (a essa forma de listar os dados se denomina Série Simples de dados).

**Tabela 3.3** Período de incubação da rubéola observado em 11 crianças, Lugar X, 2000.

Crianças observadas	Período de incubação (em dias)
criança 1	19 dias
criança 2	16 dias
criança 3	37 dias
criança 4	15 dias
criança 5	16 dias
criança 6	32 dias
criança 7	15 dias
criança 8	16 dias
criança 9	20 dias
criança 10	16 dias
criança 11	15 dias

O valor mais frequente, ou seja o que mais se repete, é denominado **moda**. No exemplo, observamos que os mais frequentes foram os períodos de incubação de 16 dias (4 vezes). Esse valor pode ser utilizado para representar o período de incubação nesse grupo; por exemplo, “o período de incubação de rubéola mais frequente é de 16 dias”.

Ordenando os valores de forma ascendente, temos, no exemplo:

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11)  
 15 15 15 16 16 **16** 16 19 20 32 37

Observaremos que o sexto valor, ou seja, o que se encontra no meio da série é o de 16 dias. Esse valor que ocupa a posição central de uma série ascendente ou descendente é denominado **mediana** e divide a série em 50% das observações acima e 50% abaixo dela. Essa medida também é útil e pode ser empregada para representar a tendência central do conjunto de dados, principalmente quando não é conveniente usar a média devido à presença de valores extremos ou quando a distribuição é assimétrica, como na curva epidêmica.

Se em vez de um número ímpar de observações, tivéssemos um número par de valores, teríamos, por exemplo:

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10)  
 15 15 15 16 **17 18** 19 20 32 37

Pode-se constatar que nessa série não há um valor central. Para calcular a mediana, somam-se os dois valores centrais (no caso, 17 e 18) e se divide o resultado por dois:

$$\frac{17+18}{2} = \frac{35}{2} = 17,5 \text{ dias}$$

A **média** ( $\bar{x}$ ) também é muito útil e se obtém somando os valores de todas as observações e dividindo o resultado entre o número de observações. Em nosso exemplo, é a soma de todos os valores dos períodos de incubação dividida pelo número de crianças observadas.

$$\frac{15+15+15+16+16+16+16+19+20+32+37}{11} = \frac{217}{11} = 19,7 \text{ dias}$$

A média de 19,7 dias é maior que os valores da moda e da mediana pois, como considera os valores de todos os casos, se vê afetado pela influência dos casos com 32 e 37 dias de incubação, que são valores extremos.

Em muitas ocasiões, os dados estão disponíveis como distribuição de frequências, em cujo caso a série se representaria da seguinte forma:

**Tabela 3.3** Período de incubação da rubéola observado em 11 crianças, Lugar X, 2000.

Período de incubação (dias)	Casos (número)
15	3
16	4
19	1
20	1
32	1
37	1

O cálculo da média desses dados seria feita assim:

$$3(15)+4(16)+1(19)+1(20)+1(32)+1(37) = 45+64+19+20+32+37 = 217/11 = 19,7 \text{ dias}$$

As medidas de tendência central são de grande utilidade também para comparar grupos de valores. Por exemplo, das pessoas que participaram em um passeio, um grupo adoeceu depois do almoço e outro grupo não apresentou nenhum sintoma. As idades das pessoas nos dois grupos foram as seguintes:

- doentes: 8, 12, 17, 7, 9, 11, 6, 3 e 13
- sadios: 19, 33, 7, 26, 21, 36, 33 e 24

As médias aritméticas calculadas foram:

$$\text{doentes} = \frac{8 + 12 + 17 + 7 + 9 + 11 + 6 + 3 + 13}{9} = 10 \text{ anos}$$

$$\text{sadios} = \frac{19 + 33 + 7 + 26 + 21 + 36 + 33 + 24}{8} = 25 \text{ anos}$$

Portanto, a doença atingiu mais as crianças do que os adultos. As crianças eram mais suscetíveis ou ficaram mais expostas (consumiram maior quantidade de alimento contaminado).

### Séries agrupadas

Quando se têm muitos dados, é necessário agrupá-los; para isso, foram construídos intervalos, que podem conter igual ou diferente número de unidades, e a eles são atribuídos os dados observados. Temos então uma série de dados agrupados como na Tabela 3.4a.

**Tabela 3.4a** Casos de suicídio por idade. Lugar X, 1995-2000

Faixa etária (anos)	Casos ( <i>f</i> )
10 a 14	37
15 a 19	176
20 a 29	693
30 a 39	659
40 a 49	784
50 a 59	1.103
60 a 69	1.005
Total	4.457

Quando nossos dados sobre uma variável contínua estão agrupados (categorizados), também podemos calcular uma média e mediana aproximadas. Para estimar a média, deve-se construir uma coluna com os pontos médios (*x*) de cada intervalo de classe da

variável e outra ( $fx$ ), resultado da multiplicação entre o valor de cada ponto médio ( $x$ ) e o número de casos ( $f$ ) do intervalo correspondente (Quadro 3.4b). A soma desses produtos ( $\sum fx$ ) dividida entre a soma de casos ( $\sum f$ ) nos dá uma aceitável aproximação da média.

Para calcular o ponto médio de cada intervalo de classe da variável, deve-se obter a média do intervalo, isso é, soma-se o limite inferior e superior do intervalo e se divide por dois; por exemplo, na Tabela 3.4a, o primeiro intervalo é de 10 a 14 anos  $((10+14,9)/2=12,5)$ . Note que utilizamos 14,9 por ser o limite superior real do intervalo. Os pontos médios dos seguintes intervalos de classe são calculados de modo similar.

**Tabela 3.4b** Casos de suicídio por idade. Lugar X, 1995-2000

Faixa etária (anos)	Casos ( $f$ )	Ponto médio ( $x$ )	$fx$	Casos acumulados
10 a 14	37	12,5	462,5	37
15 a 19	176	17,5	3.080,0	213
20 a 29	693	25,0	17.325,0	906
30 a 39	659	35,0	23.065,0	1.565
40 a 49	784	45,0	35.280,0	2.349
50 a 59	1.103	55,0	60.665,0	3.452
60 a 69	1.005	65,0	65.325,0	4.457
Total	$\sum f$ 4.457		$\sum fx$ 205.202,5	

Nesse exemplo, a aproximação à média seria:

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{205.202,5}{4.457} \approx 46,0 \text{ anos}$$

Se utilizássemos no cálculo os 4.457 casos, um a um, obteríamos uma média de 45,7 anos de idade.

Também podemos estimar a mediana a partir de dados agrupados em uma tabela de frequências. Para isso, teremos que construir uma coluna de “casos acumulados” como a que foi apresentada na Tabela 3.4b, que se obtém a partir da coluna de “casos”. A primeira célula corresponde aos casos de 10 a 14 anos (37), a segunda se obtém somando a esses 37 os casos de 14 a 19 anos (176) e assim sucessivamente, até completar a última célula, cujo valor tem de coincidir com o total de casos (4.457). Uma vez construída a coluna de frequências acumuladas, podemos aproximar a mediana da idade conforme o seguinte processo de cálculo: primeiro, localizar o intervalo de classe que contém a posição da mediana (PM), isso é:

$$PM = \frac{(\sum f) - 1}{2} = \frac{4.457 - 1}{2} \approx 2.228$$

Na coluna de casos acumulados o caso nº 2.228, está situado no intervalo de idade de 40 a 49 anos.

Depois de obter a posição da mediana, estima-se a mediana por interpolação; isso é:

$$mediana = Li + \left[ \frac{PM - fIPM}{fPM} \right] (AIPM)$$

onde:

- $Li$  = Limite inferior do intervalo da posição da mediana.
- $PM$  = Posição da mediana.
- $fIPM$  = Frequência acumulada do intervalo anterior à posição da mediana.
- $fPM$  = Frequência da posição da mediana.
- $AIPM$  = Amplitude do intervalo da posição da mediana.

$$mediana = 40,0 + \left[ \frac{2.228 - 1.565}{784} \right] (10) = 40,0 + 8,45 \approx 48,45 \text{ anos}$$

onde:

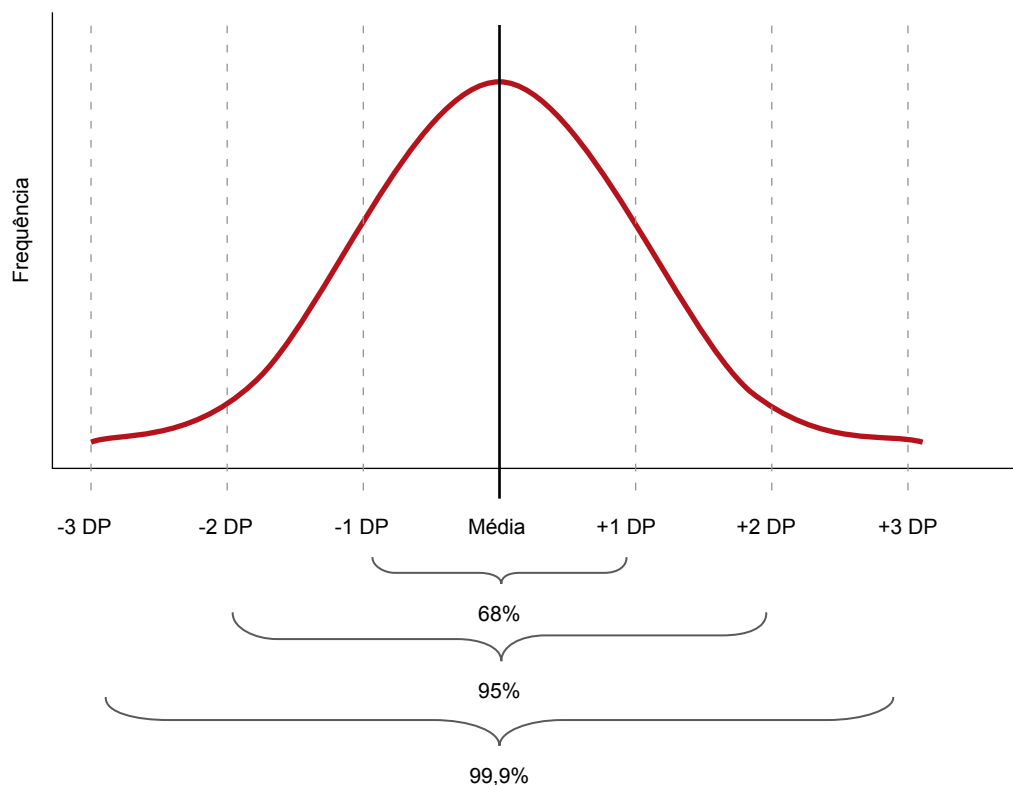
- 2.228 = caso situado no ponto médio da série
- 1.565 = casos acumulados no intervalo anterior ao que contém o caso nº 2.228
- 784 = casos do intervalo que contém a mediana
- 10 = amplitude do intervalo

Se calculássemos a mediana de idade dessa série a partir dos 4.457 casos de suicídio *individualmente*, essa seria de 48 anos de idade.

## Medidas de dispersão

Para as variáveis quantitativas contínuas, as medidas de dispersão mais usadas são basicamente três: a amplitude, a variância e o desvio-padrão. Essas medidas representam a dispersão ou variabilidade dos dados contínuos. A **amplitude** é a diferença entre o valor máximo e o valor mínimo de uma série de dados. A **variância** ( $s^2$ ), que mede o desvio médio dos valores individuais em relação à média, é o quociente entre a soma dos quadrados da diferença entre cada valor e a média, e o número de valores observados menos 1. O **desvio-padrão** (DP) é a raiz quadrada da variância. O desvio-padrão junto com a média permitem descrever a distribuição da variável. Se a variável se distribui normalmente, então o 68% de seus valores estará dentro de  $\pm 1$  desvio-padrão da média, 95% dentro de  $\pm 2$  e 99,9% dentro de  $\pm 3$  DP (Gráfico 3.7). É necessário notar que se trata de aproximações, por exemplo, a área abaixo à curva que abrange 95% não é exatamente 2 desvios, senão 1.96.

Gráfico 3.7 Distribuição normal



Vejamos um exemplo. Voltando ao Tabela 3.3, a amplitude do período de incubação da rubéola, com base nas 11 observações, é de 22 dias (isto é, 37-15 dias). A variância seria:

$$\text{variância} = \frac{(15 - 19,7)^2 + (15 - 19,7)^2 + \dots + (16 - 19,7)^2 + \dots + (32 - 19,7)^2 + (37 - 19,7)^2}{11 - 1}$$

$$\text{variância} = \frac{572,19}{10} = 57,219 \text{ dias}^2$$

e o desvio-padrão seria:

$$\text{desvio - padrão (DP)} = \sqrt{57,219} = 7,56 \text{ dias}$$

Resumindo, as estimativas dos parâmetros da distribuição do período de incubação da rubéola nas 11 crianças do exemplo seria:

$$\text{média } (\bar{x}) = 19,7 \text{ dias}$$

$$\text{desvio-padrão (DP)} = 7,6 \text{ dias}$$

Quando nossos dados sobre uma variável contínua estão agrupados, também podemos calcular a variância e o correspondente desvio-padrão aproximados. Para isso, da mesma forma que para estimar a média em dados agrupados, parte-se da coluna com os pontos médios ( $x$ ) de cada intervalo de classe. Em seguida, geramos três colunas; uma com as diferenças entre o ponto médio de cada intervalo de classe e a média de nossos dados ( $x - \bar{x}$ ), chamada desvio; outra com esse desvio elevada ao quadrado ( $(x - \bar{x})^2$ ), chamada desvio quadrático, e, finalmente, outra coluna ( $f(x - \bar{x})^2$ ), resultado da multiplicação do desvio quadrático pelo número de casos ( $f$ ) do intervalo correspondente (Tabela 3.4c). A soma de todos esses produtos ( $\sum f(x - \bar{x})^2$ ) dividida entre a soma de casos ( $\sum f$ ) nos dá uma razoável aproximação à variância e, extraindo-se a sua raiz quadrada, obteremos uma estimativa aceitável do desvio-padrão de nossos dados agrupados.



**Tabela 3.4c** Casos de suicídio por idade. Lugar X, 1995-2000

Faixa etária (anos)	Casos ( $f$ )	Ponto médio ( $x$ )	Desvio ( $x - \bar{x}$ )	Desvio quadrático ( $(x - \bar{x})^2$ )	Produto $f(x - \bar{x})^2$
10 a 14	37	12,5	-33,5	1.124,97	41.623,71
15 a 19	176	17,5	-28,5	814,56	143.362,57
20 a 29	693	25,0	- 21,0	442,70	306.792,87
30 a 39	659	35,0	- 11,0	121,89	80.327,22
40 a 49	784	45,0	- 1,0	1,08	848,79
50 a 59	1.103	55,0	9,0	80,27	88.540,76
60 a 69	1.005	65,0	19,0	359,46	361.260,03
Total	$\Sigma=4.457$				$\Sigma=1.022.755,94$

média ( $\bar{x}$ ) = 46,0 anos

Nesse exemplo, a aproximação à variância ( $s^2$ ) seria:

$$s^2 = \frac{\sum f(x - \bar{x})^2}{\sum f} = \frac{1.022.763,25}{4.457} \approx 229,47$$

e o correspondente desvio-padrão (D.P., ou “S”) para os dados agrupados seria:

$$D.P. = \sqrt{229,47} \approx 15,1 \text{ anos}$$

Se utilizássemos no cálculo os 4.457 casos, um a um, obteríamos um desvio-padrão de 15,3 anos de idade.

Finalmente, outra forma útil de representar a dispersão da distribuição de uma série de dados é usando **quantis**, que são os valores que ocupam uma determinada posição em função da quantidade de partes iguais em que foi dividida uma série ordenada de dados. Se dividimos nossa série em 100 partes iguais, falamos de *percentis*; se a dividimos em 10 partes iguais, *decis*; em cinco partes, *quintis*; e em quatro, *quartis*. Como exemplo, a Tabela 3.5 apresenta nove percentis de idade correspondentes à série de casos de malária da Tabela 3.2: o percentil 25 dessa distribuição é 33 anos, que equivale a dizer que 25% dos casos tinham idade menor ou igual a 33 anos (e conseqüentemente, 75% dos casos eram de mais de 33 anos). Observe que a mediana da distribuição corresponde ao percentil 50 (ou quartil 2), que no exemplo é 38 anos de idade.

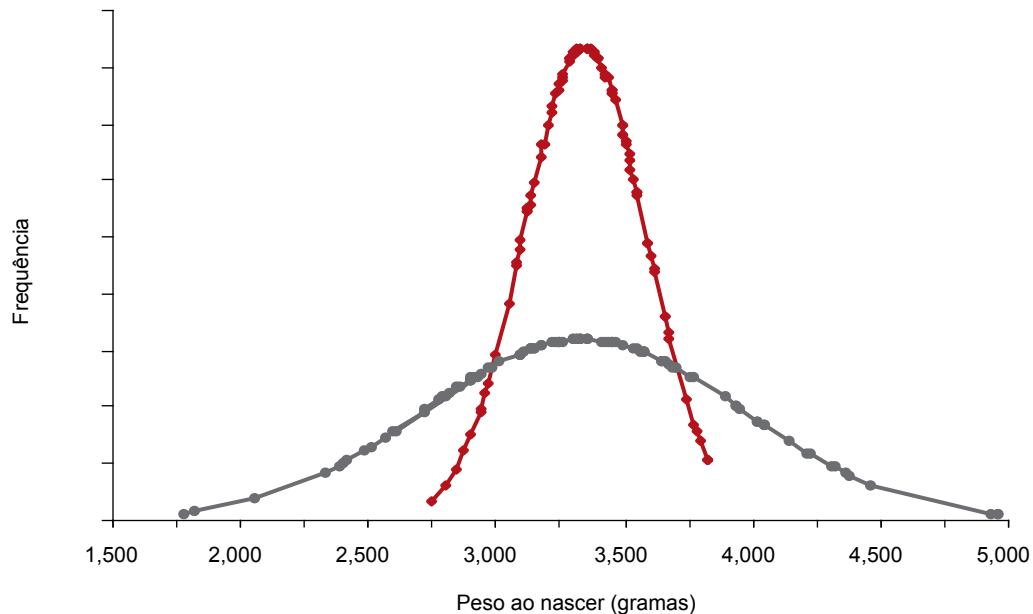
**Tabela 3.5** Percentis das idades dos casos de malária - Ilha do Caribe, 2001

Percentis	Idade
1%	24
5%	27
10%	29
25%	33
50%	38
75%	44
90%	50
95%	54
99%	63

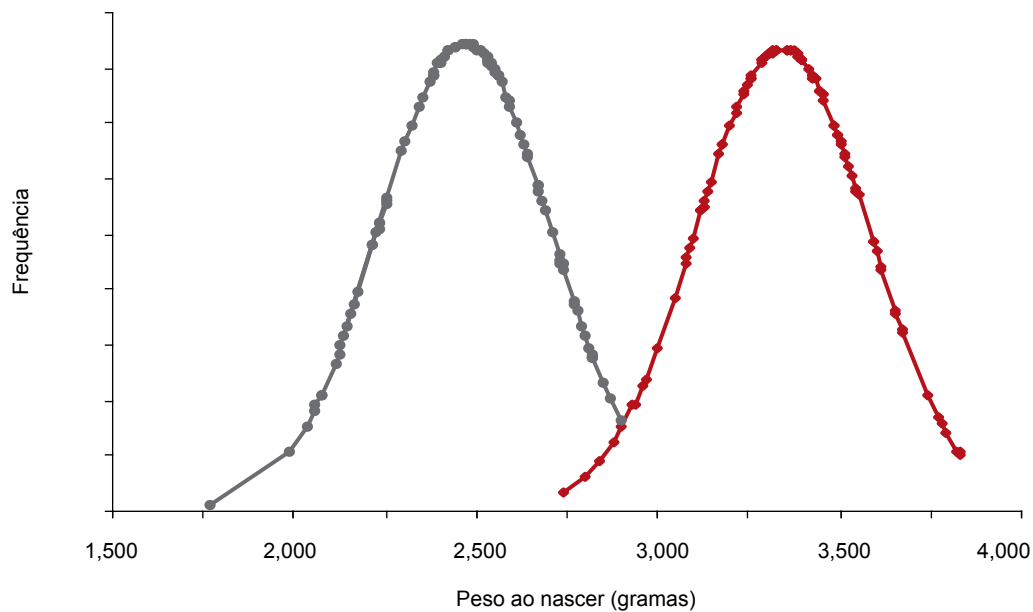
A diferença entre os percentis 25 e 75 (ou quartis 1 e 3, respectivamente) é denominada de **amplitude interquartil**, que é outra medida específica da dispersão de uma distribuição. A amplitude interquartil, que inclui o 50% central de valores na série de dados (no exemplo, essa amplitude é de 11 anos), é uma medida muito aplicada na vigilância em saúde pública (Módulo 4), especialmente para a elaboração de canais endêmicos.

Como vimos, a média e o desvio-padrão definem a **distribuição normal** e, por isso, são considerados seus *parâmetros*. A média, como medida-resumo de tendência central dos dados, é um indicador-medida-resumo da precisão das observações. Por um lado, o desvio-padrão, como medida-resumo da dispersão dos dados, é um indicador da *variação* das observações. Esses dois conceitos, precisão e variação, são de grande importância para documentar a *incerteza* com que observamos os fenômenos na população e constituem os princípios básicos do processo de inferência estatística, cujo uso nos permite derivar conclusões acerca de toda a população, observando somente uma amostra da mesma. O Gráfico 3.8 ilustra esquematicamente duas situações possíveis e frequentes em relação à distribuição do peso ao nascer na população.

**Gráfico 3.8** Curvas de distribuição normal do peso ao nascer



**A.** mesma precisão, diferente variação (média igual, desvio-padrão diferente)



**B.** mesma variação, diferente precisão (média diferente, desvio-padrão igual)

Uma vez revisada a forma de apresentação dos dados, as medidas de tendência central e de dispersão dos mesmos, continuamos com o estudo das distintas distribuições ou modelos teóricos. O mais conhecido deles é a chamada Curva Normal (Gráfico 3.7), que acabamos de introduzir. Existem outros modelos teóricos que têm aplicação na Epidemiologia e Saúde Pública, mas vão além dos objetivos desta proposta. Destaca-se que a Distribuição Normal define-se por 4 elementos característicos:

1. Tem um eixo de simetria.
2. A média aritmética, a mediana e a moda coincidem no mesmo valor pelo qual passa o eixo de simetria.
3. A distância entre o eixo de simetria e os pontos de inflexão da curva equivalem ao desvio-padrão.
4. É assintótica ao eixo “x” (abscissas), isto é, seus extremos se aproximam cada vez mais do eixo horizontal, sem nunca tocá-lo.

## Medidas de frequência

Como revisamos no Módulo 2 do MOPECE, o enfoque epidemiológico parte da **observação** de fatos na população. Esses fatos são a presença de doenças, a exposição a determinados fatores, ou outros atributos ou eventos de interesse. São “de interesse” porque consideramos que, eventualmente, agindo sobre os mesmos, poderemos modificar de forma positiva a realidade observada. O segundo aspecto desse foco epidemiológico é a **quantificação**, isso é, atribuir números aos fatos e para isso contamos com procedimentos estatísticos básicos, como os que foram descritos até agora nesse Módulo. Assim, medir a **frequência** ou *ocorrência* de tais fatos na população é uma tarefa fundamental. Entretanto, como este processo não está pautado numa certeza absoluta, para medir a frequência das doenças na população, o epidemiologista recorre a três conceitos muito importantes: probabilidade, risco e taxa.

Na sua interpretação mais simples, a **probabilidade** é uma medida de ocorrência de um fato incerto, ou seja, o número de vezes que se espera que ocorra um evento no futuro. O **risco à saúde** é uma medida que demonstra a *probabilidade* de produzir um efeito ou dano à saúde dentro de um período de tempo estabelecido. A **taxa** é uma medida da velocidade de mudança de um estado ao outro (por exemplo, de sadio a doente) e, portanto, também expressa um *risco*. No entanto, diferente do risco, a taxa implica a experiência da exposição da população nas unidades efetivas de tempo-pessoa.

**Probabilidade:** Número de eventos que ocorrem dentro de um número de eventos possíveis.

**Risco à saúde:** probabilidade de experimentar um efeito adverso ou dano em um tempo determinado.

**Taxa:** medida da velocidade de mudança de um fenômeno dinâmico por unidade de população e de tempo (tempo-pessoa de exposição).

A probabilidade é mensurável ou quantificável e seu valor numérico situa-se entre 0 e 1 (sendo 1 a certeza absoluta). Por conveniência, pode-se expressar como um número decimal, uma fração ou uma porcentagem.

O risco também é mensurável e é obtido a partir de um quociente ou fração. No sentido estrito, do mesmo modo que a probabilidade, seu valor se situaria entre 0 e 1. Costuma ser expresso como um valor da fração multiplicado por uma constante.

A taxa é a medida clássica da epidemiologia. Tem três componentes básicos:

- um *numerador*, que corresponde ao número de indivíduos que experimentam o evento de interesse (por exemplo, óbito, doença, nascimento, internação hospitalar);
- um *denominador*, que corresponde ao número total de indivíduos na população de expostos ou em risco de apresentar o evento; e,
- um *período de tempo específico*, durante o qual foi observada a frequência do evento de interesse e a população que ficou exposta efetivamente.

A partir disso, afirma-se que a taxa pode assumir um valor de 0 a infinito e se expressar em número de eventos que ocorrem em um tempo-pessoa de exposição, que é a característica fundamental da taxa.

Como se pode observar, a taxa combina em uma só expressão as três dimensões básicas da análise epidemiológica: pessoa, lugar e tempo. Aplicada corretamente, a taxa não somente é uma medida de frequência da doença na população, como também é uma medida do risco de adoecer na população. Dessa forma, ao fazer uso das taxas, poderíamos comparar a frequência e o risco de doenças em diferentes populações, diferentes subgrupos da mesma população ou diferentes períodos de tempo na mesma população.

## Prevalência e incidência

As clássicas medidas de frequência de doença são duas: prevalência e incidência, que veremos nesse Módulo. Mais especificamente, revisaremos a **incidência acumulada**, e a **taxa de incidência**. Ainda que seja de uso comum, o termo “taxa de prevalência” deve ser evitado pois não representa uma taxa, logo não é uma medida dinâmica; uma denominação mais apropriada é **proporção de prevalência**, que também será revisada.

## Medidas de morbidade

Na quantificação das condições de saúde e doença na população, é imprescindível o uso de indicadores que facilitem sua descrição e análise. Considera-se que esses estimadores são *convencionais*, pois o consenso científico-técnico e a prática confirmam sua utilidade como ferramentas relevantes para a interpretação do perfil de doença na população, com fins comparativos.

Para poder responder às necessidades das populações de forma adequada, os serviços de saúde devem efetuar sistematicamente medidas com o objetivo de determinar a frequência das condições de saúde relevantes na comunidade. Por exemplo, reconhecer que existem 700 pessoas na comunidade com diagnóstico de hanseníase é uma informação essencial para organizar os recursos existentes e obter, desde outro nível do sistema se

fosse necessário, o apoio adicional requerido para a atenção de todos os doentes. Desse modo, a contagem dos casos de uma doença é uma medida de grande importância que serve para orientar os serviços quanto à magnitude dos recursos necessários.

A contagem de nascimentos permite estimar a quantidade de crianças com menos de 1 ano de idade que existem em uma comunidade em um determinado ano. Essa medida pode ser usada, por exemplo, para projetar a quantidade de vacinas necessárias para imunizar a referida população nesse ano. Essa informação poderá ter maior precisão se, além disso, é conhecido também o número de óbitos de menores de um ano ocorridos nessa mesma comunidade no ano considerado. Isso é, em uma população e tempo específicos, o número de crianças a serem vacinadas pode ser operacionalmente definido como o número de crianças sobreviventes ao primeiro ano de vida e pode ser estimado pela diferença entre o número de nascidos vivos e o número de óbitos de menores de 1 ano de idade.

Consequentemente, a **enumeração ou contagem** de casos de doença (nascimentos e óbitos, junto com a estimativa dos censo populacionais e suas características), constituem os dados básicos que permitem aos serviços de saúde obter um melhor conhecimento sobre as condições de saúde e doença nas populações e, portanto, obter melhor desempenho.

Por outro lado, é praticamente impossível conhecer a magnitude *real* da **morbidade** em uma população devido a múltiplos e complexos fatores, que vão desde a diversidade da percepção cultural sobre saúde e doença e a presença de casos subclínicos até a falta de acesso aos serviços de saúde e a falta de confiança na assistência médica. No entanto, é a *disponibilidade de registros de boa qualidade* que fazem possível obter estimativas epidemiologicamente válidas, que nos aproximam ao conhecimento da realidade e nos permite modificá-la positivamente.

Consideramos, como exemplo, uma área geográfica determinada em um país da América Latina. Na população, foram observados 60 casos de tuberculose no ano 1995. Uma nova contagem no ano 2000 demonstrou a existência de 80 casos na população.

**Tabela 3.6a** Casos de tuberculose em uma comunidade latino-americana por ano. 1995-2000

Ano	Nº casos
1995	60
2000	80

*Qual seria a melhor explicação da diferença observada entre os dois anos? Os fatos que poderiam explicar a diferença observada podem ser resumidos a seguir:*

- a atenção aos doentes e demais medidas de controle foram inadequadas e por isso ocorreu um aumento no número de casos;
- as medidas de busca ativa e/ou uma mudança na definição de caso de tuberculose permitiram uma melhor identificação de casos antes desconhecidos, dando a impressão de aumento da doença na população;
- certos fatores socioeconômicos, independentemente das medidas de controle, provocaram um aumento da doença (por exemplo, crise econômica, desemprego, desnutrição); e/ou,
- houve um aumento da população local, por crescimento natural ou por imigração (por exemplo, atração por oferta de emprego em áreas de desenvolvimento industrial).

Relacionando o número de casos com o total da população existente nos dois anos, teremos:

**Tabela 3.6b** Casos de tuberculose em uma população latino-americana por ano. 1995-2000

Ano	Nº de casos	Total população
1995	60	30.000
2000	80	50.000

Observamos que se, por um lado, houve um aumento de 60 para 80 no número de casos, também aumentou de 30.000 a 50.000 a população no mesmo período. Portanto, o que desejamos comparar é a diferença entre 60 casos em 30.000 pessoas e 80 casos em 50.000.

Um cálculo simples nos permitirá a comparação mais direta.

- em 1995: →  $60 / 30.000 = 0,0020$
- em 2000: →  $80 / 50.000 = 0,0016$

Com a finalidade de facilitar a comparação, expressando a proporção em números inteiros e não decimais, costuma-se multiplicar o resultado por 100, 1.000, 10.000 ou 100.000, conforme a nossa conveniência ou seguindo uma convenção preestabelecida. No exemplo, se multiplicarmos o resultado da divisão entre casos e população por 10.000 teremos:

- em 1995: → 20 casos por 10.000 habitantes
- em 2000: → 16 casos por 10.000 habitantes



Isso nos permite afirmar que houve uma diminuição da prevalência de tuberculose nesse período de tempo e nessa comunidade latino-americana.

**Tabela 3.6c** Prevalência (por 10.000 pessoas) da tuberculose em uma comunidade latino-americana

Ano	Prevalência (por 10.000hab.)
1995	20
2000	16

O cálculo realizado foi o da *proporção* de prevalência da tuberculose (na área específica, nos anos 1995 e 2000).

**Prevalência:** é a medida do número total de casos existentes, chamados casos prevalentes, de uma doença em um ponto ou período de tempo e em uma população determinada, sem distinguir se são casos novos ou não. A prevalência é um indicador da magnitude da presença de uma doença ou outro evento de saúde na população.

$$\text{Prevalência da doença A} = \frac{\text{N}^\circ \text{ de pessoas com a doença A num período}}{\text{N}^\circ \text{ total de pessoas no mesmo período}} \times \text{fator}$$

Como já foi mencionado, a prevalência não pode ser considerada uma taxa, pois não considera o início nem a duração da doença. Seu denominador não necessariamente corresponde à população *em risco*, isso é, aquela população exposta e suscetível de onde surgem os casos. No entanto, a prevalência é um indicador de grande importância na saúde pública, pois fornece uma medida do volume ou carga da doença em uma comunidade ou população em um dado momento – informação essencial no planejamento dos serviços de saúde.

Do ponto de vista epidemiológico, há um maior interesse em conhecer especificamente quantos casos *novos* de uma doença aparecem em uma população durante um período de tempo. Mais concretamente, é importante conhecer quantos casos novos surgem de uma população *que está em risco* de sofrer uma determinada doença ou dano à saúde, isso é, um indicador da velocidade de mudança do processo dinâmico de saúde e doença na população. A medida de ocorrência de casos novos de doença em uma população sob risco em um tempo determinado denomina-se **incidência**. Geralmente, a incidência nos proporciona uma idéia do risco médio que existe nos indivíduos da população de sofrer a doença, bem como avaliar a eficácia das ações de controle adotadas.

**Incidência:** é a medida do número de casos novos, chamados casos incidentes, de uma doença, originados de uma população em risco de sofrê-la, durante um período de tempo determinado. A incidência é um indicador da velocidade de ocorrência de uma doença ou outro evento de saúde na população e, conseqüentemente, é um estimador do risco absoluto de vir a padecer da mesma.

$$\text{Incidência da doença B} = \frac{\text{n}^\circ \text{ de casos novos da doença B em um período}}{\text{n}^\circ \text{ total de pessoas em risco no começo do mesmo período}} \times \text{fator}$$

Nos cálculos de incidência e de prevalência, sempre é importante deixar bem claro qual é a população e a que período de tempo se refere. Pode-se relacionar com a população inteira de uma região ou com um grupo específico que estaria exposto ao problema. Assim, o denominador de incidência de câncer de colo de útero deveria incluir unicamente mulheres e o de incidência de gonorréia a população sexualmente ativa.

Por exemplo, a incidência de gastroenterite, na Província do Sul, durante o mês de dezembro de 2001 foi de 20 por mil em crianças de 5 a 10 anos. Isso é:

$$\begin{array}{l} \text{Incidência de} \\ \text{gastroenterite em} \\ \text{crianças de 5 a 10 anos na} \\ \text{Província do Sul no mês} \\ \text{de dezembro de 2001} \end{array} = \frac{\begin{array}{l} \text{Número de crianças de 5 a 10 anos} \\ \text{que desenvolveram gastroenterite} \\ \text{no mês de dezembro de 2001} \end{array}}{\begin{array}{l} \text{População de crianças de 5 a 10} \\ \text{anos residentes na Província do Sul} \\ \text{no mês de dezembro de 2001} \end{array}} \times \text{fator}$$

É necessário destacar que as fórmulas sobre incidência que acabamos de apresentar nesse exemplo correspondem especificamente ao que se denomina **incidência acumulada**. Calculando-se como o quociente entre o número de casos novos e o tamanho da população em risco em um período de tempo, a incidência acumulada assume que *todos* os indivíduos da população em risco estiveram *efetivamente* em risco de apresentar a doença durante *todo* o período de tempo observado. Intuitivamente, sabemos que isso raramente acontece, a princípio porque no momento que apresenta a doença, a pessoa deixa de estar em risco (por exemplo, deixa de “pertencer ao denominador”), pois passa a ser um caso (por exemplo, “passa ao numerador”).

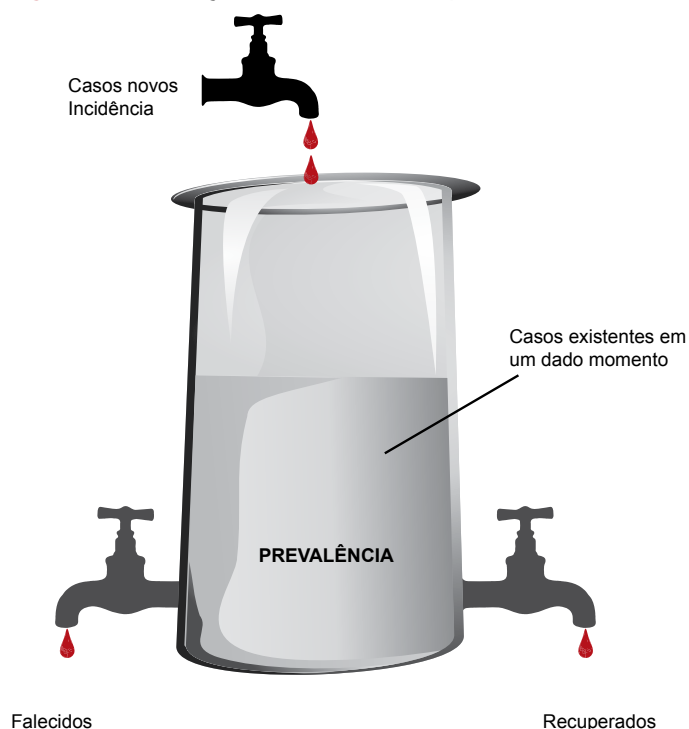
Um modo mais preciso de representar o risco de passar do estado sadio ao estado doente, ou seja estimar a incidência, é proporcionada pela investigação epidemiológica de fatores de risco causais em *coortes* de população, como veremos no Módulo 5. Em tais condições, é possível “seguir” a população, medindo com exatidão o tempo que cada indivíduo esteve em risco de adoecer e obter o número total de “tempo-pessoa” de observação (por exemplo, anos-pessoa, dias-pessoa). Dividindo agora o número de casos

novos entre o número total de anos-pessoa observados, teremos a **taxa de incidência** propriamente dita, chamada também *densidade de incidência ou força de morbidade* (Miettinen, 1975).

Em termos práticos, quando calculamos a incidência de uma doença na população, frequentemente, empregamos a incidência acumulada, pois o cálculo da taxa de incidência propriamente dita costuma estar restrito ao contexto de uma pesquisa epidemiológica. Uma forma de estimar a taxa de incidência quando não se conta com os anos-pessoa consiste em tomar como denominador a população na metade do período.

Tanto a prevalência como a incidência são medidas de morbidade (doença) na população, mas diferem em que a prevalência mede o número de pessoas que têm a doença em um dado momento (por exemplo, sua magnitude) e a incidência mede os casos novos que se apresentam em um período determinado de tempo (por exemplo, a sua velocidade). As relações entre incidência e prevalência podem ser apreciadas na seguinte figura:

**Figura 3.1** Relação entre incidência e prevalência



Observe que, se aumenta a incidência (aparecem mais casos novos) e o número de óbitos e recuperados se mantém sem alterações, aumentará a prevalência. Caso aumente a mortalidade ou mais gente se recupere e a incidência não se altera, a prevalência dimi-

nuirá. Suponhamos que é introduzido um novo teste que detecta a presença de doença precocemente no período sub-clínico; o resultado prático será um aumento na incidência, na duração da doença e também na prevalência. Por outro lado, se for introduzido um medicamento que adia ou evita a mortalidade prematura mas não cura definitivamente, o resultado também será um aumento na prevalência.

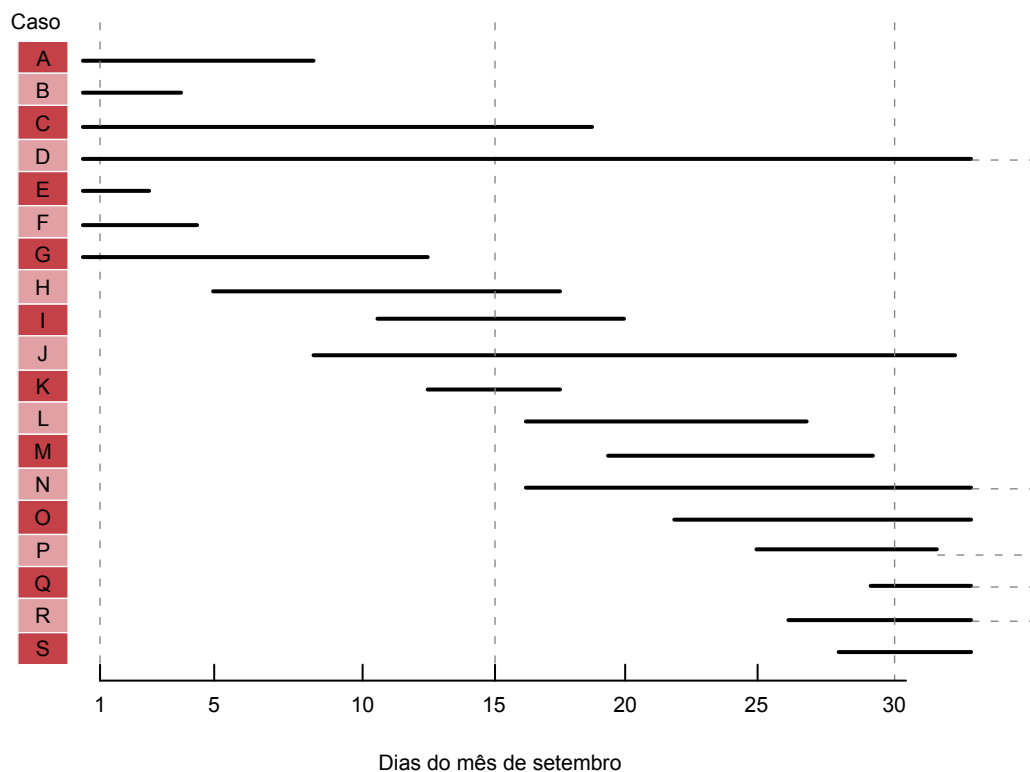
A duração da doença (D) também influencia nessa dinâmica e, em termos gerais, pode-se afirmar que em situação de equilíbrio, a prevalência (P) é o produto da incidência (I) pela duração da doença ( $P=I \times D$ ). Essa dinâmica entre incidência, prevalência e duração da doença tem importantes implicações para o controle de doenças na população, como veremos na unidade 6 do MOPECE.



### Exercício 3.1

Na Figura 3.2, cada linha representa um caso de doença respiratória (pneumonia) e a duração em dias de cada caso, que foram apresentados durante o mês de setembro.

**Figura 3.2** Casos de doença respiratória no mês de setembro



Responda às seguintes perguntas:

1. Qual é o número de casos incidentes da doença no mês de setembro?
2. Qual é o número de casos prevalentes no dia 15 de setembro?

### Outras medidas de Morbidade

Então: que medida de incidência se obtém quando é investigada uma epidemia? O numerador é o número de casos novos de doença e o denominador é o total de pessoas expostas ao fator de risco ou agente causal. O tempo, geralmente, é tratado de forma implícita, já que a maioria dos casos costuma ocorrer durante horas, dias ou semanas, conforme a doença seja tratada. A taxa de incidência que se obtém em uma situação de surto ou epidemia é denominada **taxa de ataque** da doença e é expressa usualmente em percentagem. Consideremos, como exemplo, uma situação em que 96 pessoas foram expostas a um agente (por exemplo, *Pseudomonas aeruginosa* contaminando a equipe cirúrgica), das quais 26 adoeceram em um período curto de tempo. A taxa de ataque é obtida da seguinte forma:

$$\text{taxa de ataque} = \frac{26 \text{ doentes}}{96 \text{ expostos}} \times 100 = 27,1 \%$$

Existe também a chamada **taxa de ataque secundário**, que mede o contágio de uma doença de transmissão pessoa a pessoa e, portanto, é de utilidade para avaliar a efetividade das medidas de controle de um surto, como veremos no Módulo 5 do MOPECE. A taxa de ataque secundário expressa o número de casos de uma doença que aparecem dentro do período de incubação entre os contatos suscetíveis expostos a um caso primário ou índice, em relação com o número total de contatos suscetíveis expostos. Calcula-se da seguinte maneira:

$$\text{taxa de ataque secundário} = \frac{\text{número de casos secundários}}{\text{número de contatos suscetíveis expostos}} \times 100$$

### Medidas de Mortalidade

A taxa de mortalidade estima o risco absoluto de morrer.

$$\text{taxa de mortalidade geral} = \frac{\text{número de óbitos}}{\text{população total}} \times 1.000$$

As taxas de mortalidade podem referir-se a toda a população de um país ou território ou restringir-se a uma comunidade, instituição ou a uma amostra populacional e podem, também, ser calculadas para grupos específicos de população, conforme sexo, idade, grupos de doenças ou outras características relevantes (em cujo caso constituem taxas específicas).

Existem alguns indicadores de mortalidade referidos a grupos populacionais específicos e aos que convencionalmente são chamados “taxas”. Particularmente, é necessário lem-

brar dois deles, de especial importância na saúde pública; trata-se da “taxa” de mortalidade infantil e da razão de mortalidade materna.

$$\text{taxa de mortalidade infantil} = \frac{\text{número de óbitos em menores de 1 ano de idade}}{\text{número de nascidos vivos}} \times 1.000$$

$$\text{razão de mortalidade materna} = \frac{\text{número de óbitos maternos}}{\text{número de nascidos vivos}} \times 100.000$$

É possível constatar que para o cálculo dessas “taxas” emprega-se o número de nascidos vivos como denominador e, portanto no sentido estrito não é uma taxa já que o denominador não é tempo. Por convenção, é utilizado o número de nascidos vivos como denominador desses indicadores porque se considera que é um dado mais viável de obter que os requeridos para construir as respectivas taxas, a saber: no primeiro caso, o total de crianças menores de um ano que devem existir na mesma população e ano onde foi feita a recontagem de óbitos; no segundo caso, o número total de mulheres que estiveram grávidas e que deram à luz na mesma população e ano.

Em algumas ocasiões, suspeita-se que uma doença determinada está causando um elevado e inusitado número de óbitos. É interessante conhecer quantas das pessoas doentes morrem, isso é, a *proporção* de casos fatais entre o total dos casos. Esse cálculo é chamado **taxa de letalidade** e é particularmente importante para avaliar a severidade de uma epidemia.

$$\text{taxa de letalidade da doença A} = \frac{\text{número de óbitos por doença A}}{\text{total de casos da doença A}} \times 100$$

Enquanto a mortalidade se refere aos óbitos entre a população (sadia ou doente), a letalidade somente se refere aos óbitos entre a população doente. Por exemplo, se em um distrito de 30.000 habitantes ocorreram 200 casos de febre tifóide com 6 óbitos, em um ano determinado, podemos afirmar que, nesse distrito e nesse ano, a mortalidade por febre tifóide foi 2 por 10.000 e sua letalidade 3%, isso é:

$$\text{mortalidade por febre tifóide} = \frac{6 \text{ óbitos}}{30.000 \text{ habitantes}} \times 10.000 = 2 \times 10.000$$

$$\text{letalidade por febre tifóide} = \frac{6 \text{ óbitos}}{200 \text{ casos}} \times 100 = 3 \times 100$$

### Distribuição proporcional

A distribuição proporcional é uma forma simples de expressar a quantidade de casos ou óbitos segundo alguma característica de interesse, como sexo, idade ou causa específica, como porcentagem do total de casos ou mortes observados. A distribuição proporcional corresponde a uma distribuição de frequências relativas simples, como foi visto no início desse Módulo. Ao contrário das taxas, a distribuição proporcional não mede o risco de adoecer ou morrer, mas somente indica como estão distribuídos os casos entre as pessoas atingidas.

Vejam os seguintes exemplos. Durante um surto de tifo exantemático ocorrido em uma comunidade andina, adoeceram 38 pessoas, sendo 24 homens e 14 mulheres. Desconhece-se o número total de homens e mulheres em dita população.

**Tabela 3.7** Distribuição proporcional de casos de tifo exantemático por sexo

Sexo	Número de casos	Distribuição proporcional
Masculino	24	63,2
Feminino	14	36,8
Total	38	100,0

Isso é:

$$\text{proporção de casos no sexo masculino: } \frac{24}{38} \times 100 = 63,2 \%$$

$$\text{proporção de casos no sexo feminino: } \frac{14}{38} \times 100 = 36,8 \%$$

Essa informação nos permite afirmar que a doença na comunidade atingiu mais aos homens que às mulheres (por exemplo, “de cada 10 doentes, pelo menos 6 eram homens”; ou, “existiram 1,7 casos em homens para cada caso em mulheres”). No entanto, essa informação não nos permite afirmar que os homens tiveram maior risco de adoecer.

### Mortalidade proporcional

Quando a distribuição se refere a dados de mortalidade, então falamos especificamente de **mortalidade proporcional**, uma medida de mortalidade útil para descrever o perfil e as contribuições de causas de morte específicas à mortalidade geral de um lugar e período determinados, isso é:



$$\text{mortalidade proporcional} = \frac{\text{número de óbitos por causa específica em um ano}}{\text{número total de óbitos no mesmo ano}} \times 100$$

As medidas de prevalência, incidência e mortalidade, tomadas em conjunto, permitem descrever **perfis de saúde** da população, úteis para a **análise da situação de saúde**, a primeira função essencial da saúde pública. Essas medidas básicas permitem comparar a importância relativa de riscos, doenças e danos à saúde nas populações entre diferentes períodos de tempo e entre localidades distintas, facilitando a identificação de necessidades e prioridades na saúde e orientando a gestão sanitária. No entanto, quando comparamos duas ou mais populações devemos ter sempre presente que existem outras características próprias de cada população que poderiam explicar as diferenças que observamos. Por exemplo, a diferença entre as taxas de mortalidade de duas populações pode estar distorcida pelas próprias diferenças na distribuição de idade ou na distribuição dos óbitos em distintas idades em cada população e, portanto, levar-nos a conclusões erradas. Isso pode ser evitado com o uso de técnica de *padronização ou ajuste de taxas*.



### Exercício 3.2

Os seguintes dados correspondem a duas localidades latino-americanas, as que denominaremos A e B, em um ano determinado.

**Tabela 3.8** População e mortalidade em duas localidades latino-americanas

Dados básicos	Localidade	
	A	B
População total	10.320	76.311
Número de nascidos vivos	850	1.226
População menor de 5 anos	3.350	6.901
Óbitos em menores de 1 ano	105	10
Óbitos em menores de 1 ano por infecções	32	4
Óbitos em menores de 5 anos	161	12
Óbitos em menores de 5 anos por diarreia	36	0

Calcule:

Indicadores de Saúde	Localidade	
	A	B
a) a taxa de mortalidade em menores de 5 anos (por 1.000)		
b) a taxa de mortalidade infantil (por 1.000)		
c) a taxa de mortalidade por diarreia em menores de 5 anos (por 1.000)		
d) a taxa de mortalidade por infecções em menores de 1 ano (por 1.000)		
e) O que se observa ao comparar as localidades A e B?		



### Exercício 3.3

**Pergunta 1.** Que denominador deve ser usado para o cálculo das seguintes medidas?

a) A incidência acumulada de dengue hemorrágico no país “X” em 2000.

---

b) A incidência acumulada de pneumonia nos escolares da Província “Y” em 2001.

---

**Pergunta 2.** Em uma ilha do Caribe a letalidade por febre tifóide é de 5%. Durante o ano 2000, ocorreram 40 óbitos por essa doença. Quantos casos de febre tifóide ocorreram nessa comunidade insular em dito ano?

---

---

---

**Pergunta 3.** Uma medida de incidência, usualmente expressa em porcentagem e que se refere a populações específicas, em períodos de tempo limitados, como por exemplo, epidemias, é denominada:

- a) Prevalência.
- b) Taxa ajustada.
- c) Taxa de letalidade.
- d) Taxa de ataque.
- e) Taxa de mortalidade.

**Pergunta 4.** Analise o seguinte quadro e de acordo com a informação apresentada, resolva os pontos considerados a seguir:

**Tabela 3.9** Óbitos por grupo de idade e população

Faixa etária (anos)	Óbitos (número)	População
Menores de 1	285	12.681
1 – 4	251	49.002
5 – 24	274	201.820
25 – 44	408	116.538
45 – 64	576	51.356
65 e mais	1.076	18.603
Total	2.870	450.000

a) A taxa de mortalidade (por 100.000) para o grupo de idade de 1 a 4 anos

---

b) A taxa de mortalidade (por 100.000) para o grupo de idade de 65 e mais anos.

---

c) A taxa de mortalidade (por 100.000) para o grupo de 25 a 44 anos.

---

d) Comente os resultados em relação à magnitude da mortalidade nos grupos de idade. Considere que esses resultados são típicos de qualquer população.

---

---

---

**Pergunta 5.** Vinte e seis casos de tuberculose foram diagnosticados na Cidade Alta entre 1º de janeiro e 30 de junho de 2000. O total de casos de tuberculose ativos em 30 de junho era de 264. A população de Cidade Alta era de 183.000 habitantes.

a) Qual foi a incidência de tuberculose na Cidade Alta no período entre 1º de janeiro e 30 de junho?

- (a) 7,6 casos novos por 100.000 habitantes.
- (b) 14,2 casos novos por 100.000 habitantes.
- (c) 27,3 casos novos por 10.000 habitantes.
- (d) 78,7 casos novos por 100.000 habitantes.
- (e) 144,3 casos novos por 10.000 habitantes.

b) Qual foi a prevalência de tuberculose na Cidade Alta em 30 de junho de 2000?

- (a) 14,2 casos por 100.000 habitantes.
- (b) 144,3 casos por 100.000 habitantes.
- (c) 158,5 casos por 10.000 habitantes.
- (d) 290,0 casos por 10.000 habitantes.
- (e) 85,2 casos por 100.000 habitantes.

c) Descreva e explique as diferenças nos valores obtidos sobre prevalência e incidência de tuberculose encontradas e indique o uso adequado de cada uma dessas medidas epidemiológicas. Existe semelhança dessa morbidade por tuberculose com a de sua área de saúde?

---

---

---

**Pergunta 6.** No terceiro domingo de fevereiro de 2001, comemorou-se, em uma comunidade de 462 habitantes, um baile de carnaval ao qual assistiram 287 pessoas. Nas duas últimas semanas de março, o centro de saúde local atendeu a 79 pessoas que foram atendidas por prurido intenso e erupção cutânea papulovesicular. Setenta e seis delas tinham assistido ao baile. Dos 161 homens participantes desse evento social, 53 adoeceram. A maioria (68,3%) dos participantes do baile tinham entre 20 e 39 anos de idade; nesse grupo encontrava-se o 82,9% dos doentes. Calcule:

- a) A taxa de ataque nos participantes da festa.
- b) A taxa de ataque nos homens.
- c) A taxa de ataque nas mulheres.
- d) A taxa de ataque no grupo de 20 a 39 anos de idade.
- e) Que interpretação você daria a esses resultados?

**Pergunta 7.** Examine a tabela 3.10 e resolva as questões a, b e c:

**Tabela 3.10** Número de casos por grupo de idade e população

Coluna 1 faixa etária (anos)	Coluna 2 casos (N°)	Coluna 3 população	Coluna 4	Coluna 5
Menores de 5	3	48	14,3	6,3
5 – 19	6	17	28,6	35,3
20 – 39	5	23	23,8	21,7
40 e mais	7	109	33,3	6,4
Total	21	197	100,0	10,7

- a) Os números na Coluna 4 representam:
- (a) A distribuição proporcional da população por idade.
  - (b) As taxas de ataque por idade.

- (c) A distribuição proporcional dos casos por idade.
- (d) As taxas de mortalidade por idade.
- (e) Nenhuma das anteriores.

b) Os números na Coluna 5 representam:

- (a) A distribuição proporcional da população por idade.
- (b) As taxas de ataque por idade.
- (c) A distribuição proporcional dos casos por idade.
- (d) As taxas de mortalidade por idade.
- (e) Nenhuma das anteriores.

c) Comente brevemente a informação apresentada na Tabela 3.10.

---

---

---

**Pergunta 8.** O programa de Atenção Integrada às Doenças Prevalentes na Infância (AIDPI) de um centro de saúde afirmou que o tempo de duração de todos os casos de pneumonia em crianças de 5 a 9 anos de idade atendidos nos últimos dois meses foi 9, 7, 11, 9, 8, 4, 6, 12, 6, 8, 8, e 5 dias. Usando essa informação, calcule:

a) A moda

---

b) A mediana

---

c) A média

---

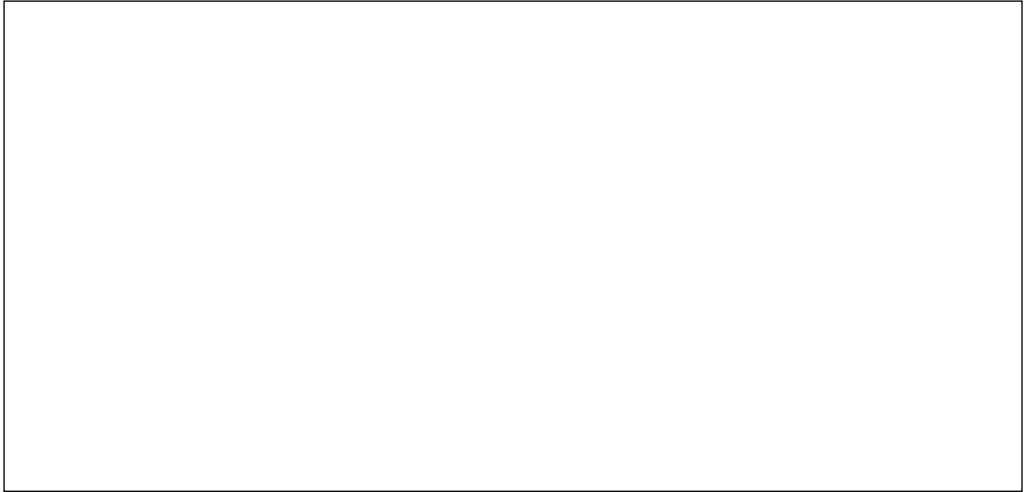
d) A amplitude

---

e) O desvio-padrão

---

f) No quadro reservado a seguir, represente a distribuição da variável de interesse, usando um gráfico apropriado.



g) Comente os resultados observados com referência a essa distribuição.

---

---





### Exercício 3.4

O nível intermediário do sistema de saúde da Nicarágua, no istmo centro-americano, compreende 17 Sistemas Locais de Atenção Integral de Saúde (SILAIS), que correspondem aos 15 departamentos e 2 regiões autônomas em que se divide geopoliticamente o país. No início de 1999, houve um surto de rubéola de alcance nacional, entrando em operação um sistema de vigilância ativa intensificada que permitiu estabelecer medidas oportunas para o controle da doença. Nesse exercício, descreve-se a situação observada no SILAIS León, localizada na região ocidental do país, e que compreende 10 municípios. A Tabela 3.11 contém dados sobre os 130 casos confirmados de rubéola ocorridos em León entre as semanas epidemiológicas (S.E.) 4 e 33 de 1999. As tabelas 3.12a, 3.12b e 3.13 apresentam dados adicionais relevantes para a caracterização da situação epidemiológica.

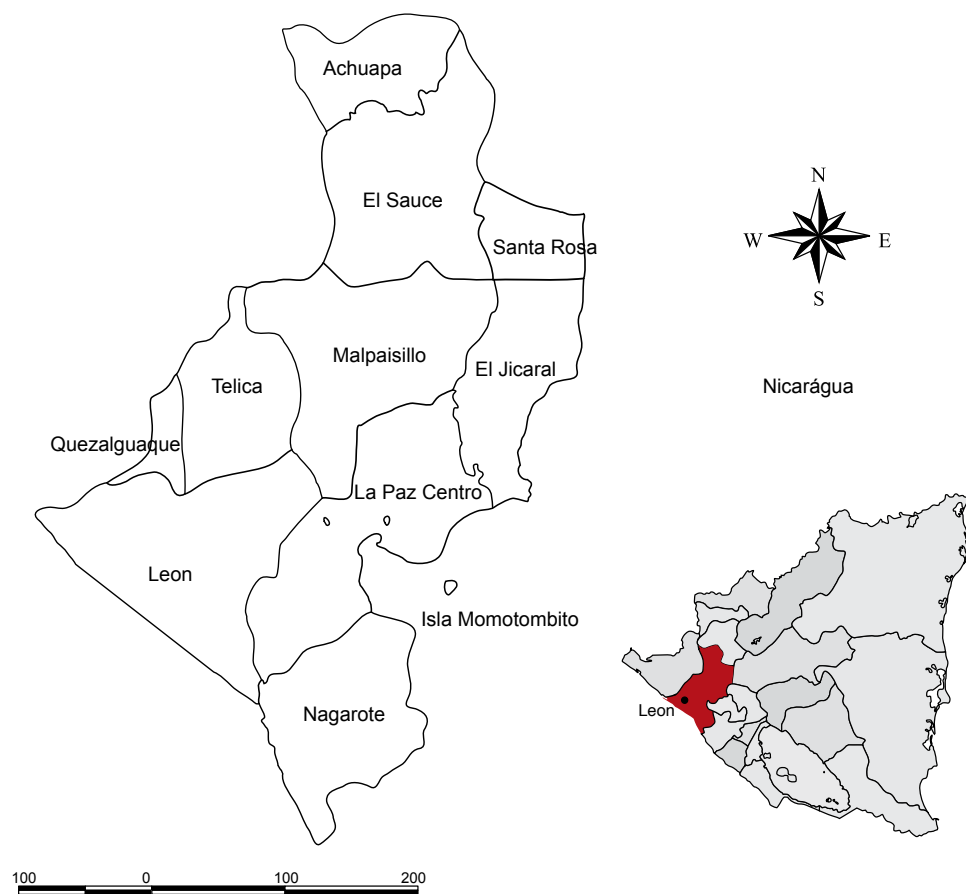
**Pergunta 1.** Quais foram as características epidemiológicas mais relevantes do surto de rubéola no SILAIS León, Nicarágua, em 1999? Caracterize a epidemia conforme:

- a) idade e sexo dos atingidos;
- b) sua distribuição geográfica; e,
- c) sua distribuição no tempo.

Para isso:

- Complete as tabelas 3.12a, 3.12b e 3.13.
- Calcule as taxas correspondentes usando as tabelas e o mapa (Figura 3.3).
- Complete o Gráfico 3.9.

**Figura 3.3** Mapa do SILAIS León, Nicarágua



Fonte: Programa Especial de Análise de Saúde (SHA) e Programa Ampliado de Imunizações (OPAS)

**Tabela 3.11** Ocorrência de casos de rubéola; SILAIS León, 1999

Nº	Sexo	Idade (anos)	Município	Data de início	S.E.*
1	F	7	León	26-Jan	4
2	F	6	León	31-Jan	5
3	F	28	León	5-Fev	5
4	F	15	León	6-Fev	5
5	F	28	León	7-Fev	6
6	F	4	León	11-Fev	6
7	F	13	León	16-Fev	7
8	F	4	León	21-Fev	8
9	F	11	Nagarote	8-Mar	10
10	M	5	León	11-Mar	10
11	F	28	León	13-Mar	10
12	F	4	León	14-Mar	11
13	F	29	León	18-Mar	11
14	M	1	León	20-Mar	11
15	M	24	León	21-Mar	12
16	F	4	León	22-Mar	12
17	F	11	León	22-Mar	12
18	M	24	Quezalguaque	23-Mar	12
19	F	9	León	25-Mar	12
20	F	3	León	29-Mar	13
21	M	4	León	29-Mar	13
22	F	8	León	4-Abr	14
23	M	3	León	6-Abr	14
24	F	34	Telica	6-Abr	14
25	M	7	León	8-Abr	14
26	F	20	Telica	8-Abr	14
27	F	6	León	9-Abr	14
28	F	7	León	9-Abr	14
29	F	10	León	10-Abr	14
30	M	13	León	11-Abr	15
31	F	9	León	12-Abr	15
32	M	7	León	13-Abr	15
33	F	6	León	14-Abr	15
34	M	12	Telica	14-Abr	15
35	F	10	León	16-Abr	15
36	F	24	León	16-Abr	15
37	F	4	León	17-Abr	15
38	M	5	León	17-Abr	15
39	F	3	Telica	18-Abr	16
40	M	3	Malpaisillo	18-Abr	16
41	M	7	León	19-Abr	16
42	F	12	Malpaisillo	19-Abr	16

Módulo 3: Medição das condições de saúde e doença na população

Nº	Sexo	Idade (anos)	Município	Data de início	S.E.*
43	F	26	León	19-Abr	16
44	F	12	León	20-Abr	16
45	M	33	Malpaisillo	20-Abr	16
46	F	11	León	21-Abr	16
47	F	17	Achuapa	21-Abr	16
48	F	5	Telica	22-Abr	16
49	F	12	León	22-Abr	16
50	F	21	Malpaisillo	22-Abr	16
51	M	6	León	23-Abr	16
52	F	10	León	23-Abr	16
53	F	10	Malpaisillo	25-Abr	17
54	M	10	Telica	26-Abr	17
55	F	30	León	26-Abr	17
56	M	11	Malpaisillo	27-Abr	17
57	F	15	León	27-Abr	17
58	M	10	Malpaisillo	1-Mai	17
59	F	9	Achuapa	2-Mai	18
60	F	10	Achuapa	2-Mai	18
61	F	6	León	3-Mai	18
62	F	26	Telica	4-Mai	18
63	F	17	Malpaisillo	7-Mai	18
64	F	19	Achuapa	8-Mai	18
65	F	16	León	10-Mai	19
66	F	11	León	11-Mai	19
67	F	22	León	11-Mai	19
68	F	29	León	14-Mai	19
69	M	22	León	16-Mai	20
70	M	9	León	17-Mai	20
71	M	7	León	19-Mai	20
72	M	5	León	20-Mai	20
73	M	22	León	20-Mai	20
74	M	4	León	21-Mai	20
75	F	5	León	21-Mai	20
76	F	20	León	22-Mai	20
77	M	1	León	24-Mai	21
78	F	7	Nagarote	24-Mai	21
79	M	1	León	26-Mai	21
80	M	4	Malpaisillo	26-Mai	21
81	M	13	León	27-Mai	21
82	M	5	León	28-Mai	21
83	F	17	León	28-Mai	21
84	F	4	León	30-Mai	22
85	F	15	León	1-Jun	22
86	F	5	Nagarote	2-Jun	22
87	F	17	Telica	4-Jun	22

Nº	Sexo	Idade (anos)	Município	Data de início	S.E.*
88	M	4	León	5-Jun	22
89	F	20	León	6-Jun	23
90	F	5	León	7-Jun	23
91	M	8	León	10-Jun	23
92	F	10	León	11-Jun	23
93	F	10	León	12-Jun	23
94	F	9	León	15-Jun	24
95	F	5	León	16-Jun	24
96	F	31	León	18-Jun	24
97	F	36	León	18-Jun	24
98	M	8	El Jicaral	20-Jun	25
99	F	12	León	21-Jun	25
100	F	8	León	22-Jun	25
101	M	10	El Jicaral	22-Jun	25
102	M	10	León	22-Jun	25
103	F	6	León	24-Jun	25
104	M	13	El Jicaral	24-Jun	25
105	M	15	El Jicaral	25-Jun	25
106	F	15	Santa Rosa	29-Jun	26
107	F	4	León	30-Jun	26
108	F	8	León	1-Jul	26
109	M	7	Santa Rosa	2-Jul	26
110	F	9	León	3-Jul	26
111	F	10	Malpaisillo	3-Jul	26
112	F	11	Malpaisillo	5-Jul	27
113	F	7	Telica	7-Jul	27
114	F	14	León	7-Jul	27
115	F	9	Nagarote	11-Jul	28
116	M	28	León	11-Jul	28
117	M	14	León	12-Jul	28
118	F	30	El Sauce	12-Jul	28
119	F	30	León	13-Jul	28
120	F	4	León	15-Jul	28
121	F	6	La Paz Centro	17-Jul	28
122	F	15	León	26-Jul	30
123	M	6	León	27-Jul	30
124	F	13	León	29-Jul	30
125	F	6	El Jicaral	30-Jul	30
126	F	7	Malpaisillo	3-Ago	31
127	F	12	Nagarote	8-Ago	32
128	F	28	Malpaisillo	8-Ago	32
129	F	7	León	11-Ago	32
130	F	26	La Paz Centro	16-Ago	33

**Tabela 3.12a** Taxa de ataque (TA) de Rubéola por 100.000 habitantes, segundo faixa etária - León, Nicarágua, 1999

Faixa etária (anos)	Casos	População	TA
0 - 4		61.932	
5 - 9		52.555	
10 -14		46.940	
15 - 19		43.285	
20 - 24		34.373	
25 - 29		28.097	
30 - 34		23.419	
35 - 39		19.318	
40 e mais		64.602	
Total		374.521	

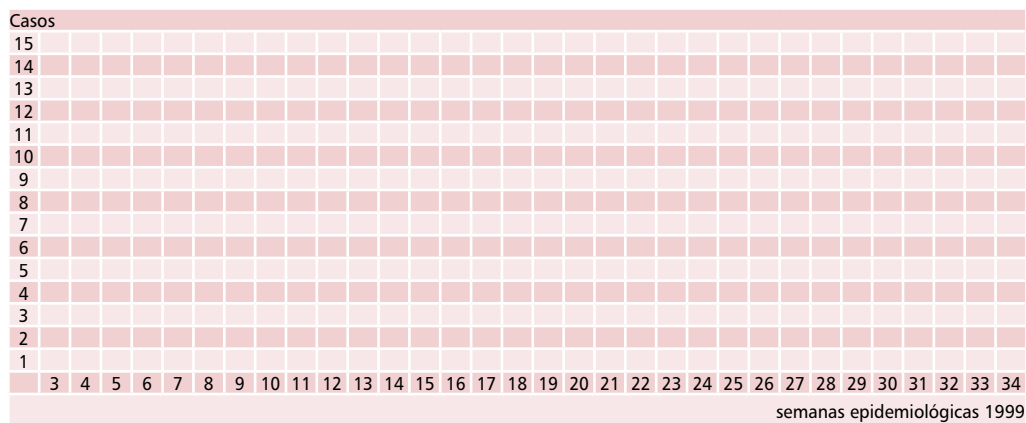
**Tabela 3.12b** Taxa de ataque (TA) de Rubéola por 100.000 habitantes, segundo faixa etária e sexo - León, Nicarágua, 1999

Faixa etária (Anos)	Homens			Mulheres		
	Casos	População	TA	Casos	População	TA
0 - 4		29.809			32.123	
5 - 9		25.539			27.016	
10 -14		22.886			24.054	
15 - 19		20.001			23.284	
20 - 24		15.770			18.603	
25 - 29		12.693			15.404	
30 - 34		10.385			13.034	
35 - 39		8.654			10.664	
40 e mais		27.347			37.255	
Total		173.084			201.437	

**Quadro 3.13** Taxa de ataque (TA) de Rubéola por 100.000 habitantes, segundo município-León, Nicarágua, 1999

Município	Casos	População	TA
Achuapa		14.681	
El Jicaral		11.174	
El Sauce		28.917	
La Paz Centro		30.627	
León		180.352	
Malpaisillo		32.025	
Nagarote		32.510	
Quezalguaque		8.710	
Santa Rosa		10.164	
Telica		25.361	
Total		374.521	

**Gráfico 3.9** Distribuição de casos de Rubéola por data de início dos sintomas. León, Nicarágua; 1999



□ = 1 caso

**Pergunta 2.** Determine quais seriam as recomendações epidemiológicas para a equipe de saúde do SILAIS de León.

---



---

## Comparação da frequência da doença e medidas de associação

Agora que já revisamos as principais medidas de frequência e distribuição dos fenômenos no tempo, espaço e pessoa, o seguinte passo do foco epidemiológico é a comparação de tais medidas. Essa comparação é a estratégia básica da análise epidemiológica e o passo fundamental para transformar os dados em informação relevante.

Em todo ato de comparação, há uma intencionalidade analítica subjacente: encontrar igualdades ou, alternativamente, encontrar diferenças. Essa capacidade de discriminar as observações que se faz da realidade é um requisito para a identificação das necessidades de saúde da população, estabelecer prioridades e, conseqüentemente, desenvolver ações em saúde dirigidas a modificá-las positivamente. Esse é o foco não igualitário que põe em prática a epidemiologia.

Um aspecto central nesse raciocínio analítico da epidemiologia é a geração de explicações, tentativas, suposições, prognósticos ou conjunturas verossímeis em relação a uma causa-efeito que envolvam incertezas, isso é, a **geração de hipóteses**. Uma hipótese é uma suposição que se faz de uma observação ou reflexão, que leva a predições refutáveis. Para a aplicação de testes estatísticos aos dados obtidos, é necessária a elaboração de duas hipóteses: a hipótese alternativa ou de investigação ( $H_A$ ) e a hipótese nula ou de não-diferença ( $H_0$ ), sobre as quais são focados os resultados da análise estatística. A obtenção de dados relevantes e a comparação racional dos mesmos é a forma de contrastar as nossas hipóteses sobre a saúde e a doença na população.



## Medidas de associação

Nesta seção revisaremos os princípios e métodos estatísticos básicos que ajudam as equipes locais de saúde a estabelecer a presença de uma **associação** entre a exposição a um fator que se considere de risco e a ocorrência de doença na população. Em termos estatísticos, se diz que duas variáveis estão associadas quando existe uma relação de dependência (geralmente estatística) entre ambas, isso é, a mudança de uma delas necessariamente acompanha a mudança na outra (covariação).

**Associação:** relação de dependência estatística entre dois ou mais eventos, características ou outras variáveis. Uma associação está presente se a probabilidade de ocorrência de um evento depende da ocorrência de outro ou outros.

Na epidemiologia, a aplicação do termo “associação” sempre implica a *intenção* de se estabelecer uma **relação de causa e efeito** entre uma exposição e uma doença ou evento na saúde. No entanto, deve-se ressaltar que uma associação pode ser casual ou “espúria” ou pode ser produzida por várias circunstâncias e, portanto, a presença de uma associação estatística não necessariamente implica uma relação causal.

Por outro lado, um **fator de risco** é um aspecto do comportamento ou estilo de vida pessoal, constituição genética ou hereditária ou exposição ambiental que, baseado na evidência epidemiológica disponível, sabe-se que está associado a condições relacionadas com a saúde, consideradas importantes de prevenir. Do ponto de vista epidemiológico, o mais importante de um fator de risco é que seja identificável, quantificável e, se possível, modificável *antes* da ocorrência do fato que o prediz. Os fatores de risco são indicadores ou marcadores do risco de adoecer da população, mas é necessário destacar que a descoberta de um fator de risco não implica necessariamente que esse seja um fator causal.

**Fator de risco:** característica ou circunstância detectável nos indivíduos ou grupos, associada com uma probabilidade incrementada de experimentar um dano ou efeito adverso à saúde. Geralmente, um fator de risco é um atributo ou exposição que incrementa a probabilidade de ocorrência de uma doença ou outro dano à saúde.

Os aspectos que foram revisados na primeira metade desse Módulo correspondem ao campo da chamada **estatística descritiva**: a *descrição* quantitativa da frequência e a distribuição dos fenômenos de saúde e doença observados na população. A utilidade de apresentar os dados resumidos em bons quadros e gráficos consiste em permitir realizar *inferências* sobre os determinantes da situação de saúde observada na população e proceder com certo nível de segurança à intervenção sobre os mesmos.

A estatística é a ciência matemática que proporciona à epidemiologia os elementos básicos para o tratamento científico da informação numérica na saúde. A análise epidemiológica usa procedimentos estatísticos desenhados para estimar a probabilidade de que uma conclusão sobre a população de referência, baseada na análise de dados de uma parte dela (amostra) ou na comparação com outra população, seja correta ou verossímil. Isso corresponde à chamada **estatística inferencial**.

Na prática, para explorar uma possível associação entre exposição e doença, são necessários três elementos: dois grupos comparáveis da população, uma medida da variável exposição para cada grupo e uma medida da variável doença em cada grupo. Geralmente, as variáveis epidemiológicas de exposição e de doença são contínuas ou discretas e suas medidas se resumirão em **médias** ou em **proporções**. A situação mais comum nos serviços de saúde é a comparação de duas proporções. Por exemplo, para avaliar a associação entre renda econômica e tuberculose, poderíamos comparar duas proporções: incidência de tuberculose entre os pobres e entre os ricos.

As medidas de associação estão baseadas nos chamados **testes de significância** (os aspectos teóricos relacionados com suas bases conceituais fogem dos propósitos desse material. Para uma breve introdução aos mesmos, veja o Apêndice ao final dessa unidade). O propósito desses testes é determinar se a presença de um fator de risco avaliado está efetivamente relacionada com a frequência da doença. Em tais condições, espera-se que a prevalência de exposição a dito fator, seja razoavelmente mais alta entre os que adoeceram ou sofreram um dano à saúde que naqueles aparentemente saudáveis.

A bioestatística possui uma grande variedade de testes de significância e outros recursos analíticos de potencial utilidade para a prática epidemiológica. Por outro lado, o desenvolvimento da informática e as possibilidades tecnológicas estão facilitando o acesso a um número cada vez maior de programas estatísticos de computação. Existem dois pacotes de programas de cálculos desenhados especificamente para a saúde que facilitam o armazenamento, processamento e análise de informação epidemiológica: Epi-Info, produzido pelo Centro de Prevenção e Controle de Doenças dos Estados Unidos (CDC) e a Organização Mundial da Saúde (OMS), e o Epi-Dat, desenvolvido pela OPAS e a Xunta de Galícia, Espanha. Longe de competir entre si, estes programas, de grande uso e de livre distribuição, oferecem processos e rotinas de manuseio e análise epidemiológica de dados que são complementares. Observa-se que o nível de *exposição* a essa tecnologia vem aumentando e, por isso, um conhecimento básico do foco estatístico e epidemiológico nas equipes locais de saúde é cada vez mais necessário para garantir o uso racional e eficiente desses recursos tecnológicos. Justamente para familiarizar as equipes locais de saúde com as aplicações da bioestatística, o MOPECE apresenta a seguir um dos testes de significância estatística de maior utilidade para a prática epidemiológica de campo.

## Comparação de duas Proporções: O Teste do Chi-quadrado

Para exemplificar o uso desse teste estatístico, nos referimos a seguir a uma situação na qual se avalia se existe uma associação estatisticamente significativa entre certos tipos de ocupação e o risco de contrair malária em uma população.

Numa localidade rural de 760 habitantes, situada em uma zona de malária, observou-se que, no último ano, a incidência acumulada de malária em camponeses foi de 88,2 por mil, enquanto que na população não camponesa foi de 55,8 por mil. De acordo com o censo local mais recente, na comunidade há 204 camponeses. Interessa saber se a ocupação camponesa está associada à malária. Os dados que resumem a situação descrita são:

	com malária	sem malária			taxa por mil	
camponês	18	186	204	$I_c =$	18/204	88,2
não-camponês	31	525	556	$I_{nc} =$	31/556	55,8
	49	711	760			

Do ponto de vista estatístico, interessa conhecer se há diferença significativa entre as duas medidas de incidência acumulada. Nesse caso pode-se aplicar um teste de significância estatística chamado Chi-Quadrado ( $\chi^2$ ), cuja fórmula é:

$$x^2 = \sum \frac{(O - E)^2}{E}$$

Na fórmula, “O” se refere ao valor observado em uma célula e “E” ao seu valor esperado. A fórmula trabalha exclusivamente com células centrais de um quadro, nesse caso uma tabela 2x2, e faz referência às frequências observadas em ditas células e às que se esperaria que ocorressem (valores esperados) se efetivamente não houvesse diferença entre as proporções comparadas. Os valores esperados para cada célula central são obtidos multiplicando seus totais marginais correspondentes e dividindo esse produto pelo grande total. Por exemplo, na primeira célula (camponês com malária), o valor observado é 18 e o valor esperado seria:  $(204 \times 49) / 760 = 13,2$ . O quadro de resultados para cada célula ficaria assim:

	com malária	sem malária	
camponês	13,2 / 18	190,8 / 186	204
não-camponês	35,8 / 31	520,2 / 525	556
	49	711	760

Substituindo na fórmula,

$$x^2 = \frac{(18 - 13,2)^2}{13,2} + \frac{(186 - 190,8)^2}{190,8} + \frac{(31 - 35,8)^2}{35,8} + \frac{(525 - 520,2)^2}{520,2} = 2,56$$

O valor de Chi-Quadrado calculado desse modo (observado) é comparado com um valor tabulado (esperado) tomado da distribuição de probabilidades teóricas. Esse valor teórico corresponde ao que seria esperado encontrar se os resultados observados ocorressem puramente por acaso. Esse valor teórico é chamado *valor crítico*: se o valor observado é maior que o valor crítico, conclui-se que a diferença observada não é devida ao acaso e se diz que é *estatisticamente significativa*.

O valor crítico indica o *nível de significância* do teste, que expressa a probabilidade de que a diferença observada tenha ocorrido por acaso, visto que, na realidade não existiam diferenças. Usualmente, esta probabilidade é estabelecida em 5% e apresenta-se como  $p < 0,05$ . O complemento dessa probabilidade se chama *nível de confiança*, em geral, 95%.

Para um nível de confiança de 95%, o valor crítico do Chi quadrado (de acordo com uma tabela de distribuição teórica) é 3,84, que corresponde ao chamado chi quadrado com um grau de liberdade, específico para tabelas 2x2. Os graus de liberdade de uma tabela se referem ao número mínimo de células centrais que são necessárias para conhecer os valores das outras células, dado que se conheçam os marginais. Nesta situação analisada, o valor observado (2,56) não ultrapassa o valor crítico (3,84), portanto, concluímos que não se pode desprezar a hipótese de não-diferença (de nulidade) Consequentemente, pode-se afirmar que com 95% de confiança estatística que a ocupação de compônes não está significativamente associada com malária ( $p > 0,05$ ).

O teste do Chi-Quadrado é amplamente usado em epidemiologia, especialmente na análise de tabelas 2x2. Por isso, desenvolveu-se a seguinte fórmula alternativa simplificada:

$$x^2 = \frac{n(ad - bc)^2}{(a + c)(b + d)(a + b)(c + d)}$$

	doente	não doente	
exposto	a	b	(a+b)
não exposto	c	d	(c+d)
	(a+c)	(b+d)	n

Essa, aplicada aos dados do nosso exemplo, proporciona resultados similares:

$$x^2 = \frac{760[(18 \times 525) - (186 \times 31)]^2}{49 \times 711 \times 204 \times 556} = 2,61 \quad (p > 0,05)$$

A decisão sobre o nível de confiança selecionado depende da situação que se analise; pelas implicações que uma decisão tenha, será preciso ter um nível de confiança maior ou menor. Para servir a distintos propósitos, existem várias alternativas. A Tabela 3.14 apresenta outros valores críticos de  $\chi^2$  para diferentes níveis de significância.

**Tabela 3.14** Valores críticos dos valores estatísticos  $X^2$  e Z em distintos níveis de significância e confiança estatísticas

Nível de significância (p)	Valores Críticos		Nível de confiança
	Chi-Quadrado*	Teste Z**	
0,001	10,83	3,29	99,999
0,01	6,63	2,58	99,99
0,05	3,84	1,96	95,0
0,10	2,71	1,64	90,0
0,20	1,64	1,28	80,0

\* com um grau de liberdade (somente para tabelas 2x2)

\*\* para comparación de dos promedios (ver Apéndice)



### Exercício 3.5

Durante o outono de 1990, em uma localidade da América do Norte, reportou-se a inusitada ocorrência de uma doença caracterizada por febre, náuseas, edema, dificuldade para respirar e, principalmente, intensas dores por contração muscular, acompanhada de uma marcante alta do número de eosinófilos, um tipo especial de glóbulos brancos que agem sobre certos processos alérgicos. Em menos de um ano, tinham sido reportados mais de 1.500 casos de Síndrome de Eosinofilia Miálgica. Os estudos epidemiológicos envolveram o consumo de triptofano, um importante aminoácido na dieta humana, como responsável da doença, especialmente associado ao consumo de um produto popular para combater a depressão, a insônia e a síndrome pré-menstrual.

**Pergunta 1.** Analise os seguintes dados e avalie se existe uma diferença estatisticamente significativa na incidência da doença entre os dois grupos estudados segundo a exposição ao triptofano. Use um nível de significância igual a 0,05. Interprete os resultados em termos de uma hipótese nula (de não diferença).

	Número	Incidência (%)
Expostos	30	66,7
Não-expostos	36	22,2

### Medidas da força de associação

Ao contrário dos chamados testes de significância estatística, úteis porque determinam a *presença* de uma associação entre duas variáveis, a epidemiologia propõe o uso de duas medidas básicas que *quantificam* a força dessa associação: o risco relativo e o OR (o OR vem do inglês *Odds Ratio* que foi traduzido como: razão de produtos cruzados, razão de chances ou oportunidade relativa). Neste Módulo, revisaremos os conceitos básicos sobre estas medidas de análise epidemiológica e no Módulo 5, as veremos no contexto da investigação epidemiológica de fatores de risco de doença, que é onde adquirem especial relevância.

### Risco Relativo

Como vimos anteriormente, a incidência de uma doença numa população e período determinados (incidência acumulada e a taxa de incidência) nos proporciona uma medida do **risco absoluto** de adoecimento nessa população.

**Risco Absoluto:** incidência da doença ou outro evento de interesse na população ou grupo populacional; quantifica a probabilidade de experimentar tal doença ou evento.

A comparação de duas medidas de incidência de uma doença, isso é, dois riscos absolutos, permite detectar um possível **excesso de risco** de um grupo *em relação ao outro*. Na epidemiologia, é de particular interesse comparar a ocorrência de doença entre um grupo exposto a um fator considerado de risco em relação a outro não exposto. Os testes estatísticos ajudam a detectar o excesso de risco entre ambos grupos; o **risco relativo** permite *quantificar* a magnitude de tal excesso e mede a força da associação entre exposição e doença. Para calcular o risco relativo de uma doença em relação a uma exposição, é necessária uma medida do risco absoluto entre os expostos a um fator de risco e uma medida do risco absoluto entre os não expostos a tal fator, isso é, a incidência nos expostos e a incidência nos não expostos. É necessário observar que estamos nos referindo à incidência *de uma doença específica e à exposição a um fator de risco específico*.

A epidemiologia dispõe de uma série de desenhos de estudos para observar quando existe uma associação entre a exposição a um fator e o desenvolvimento subsequente de uma doença. Entre esses desenhos, os estudos de *coortes* e caso-controle não só demonstram se essa associação existe senão quanto ela é forte. Os dados obtidos a partir desses estudos observacionais são apresentados usualmente em uma tabela 2x2, chamada assim pela natureza dicotômica das variáveis de exposição e de doença, como a que se mostra a seguir:

**Tabela 3.15** Tabela 2x2 (Estudo de *coortes*)

	Doente	Não doente	
Exposto	a	b	a + b
Não exposto	c	d	c + d
	a + c	b + d	a + b + c + d

Convencionalmente, as colunas da tabela 2x2 representam a presença ou ausência de doença e as linhas a presença ou ausência de exposição. Nos estudos de *coorte*, parte-se de dois grupos de indivíduos *sem a doença*, um exposto a um hipotético fator de risco e outro não exposto ao fator, em seguida se observa o desenvolvimento da doença em ambos os grupos, durante um tempo de seguimento. Dessa forma, nos estudos de *coorte*, a tabela 2x2 tem os seguintes componentes:

- a = expostos ao fator de risco que durante o segmento desenvolveram a doença.
- b = expostos ao fator de risco que durante o seguimento não desenvolveram a doença.

- c = não expostos ao fator de risco que durante o segmento desenvolveram a doença.
- d = não expostos ao fator de risco que durante o segmento não desenvolveram a doença.
- a + c = total de indivíduos que durante o segmento desenvolveram a doença.
- b + d = total de indivíduos que durante o segmento não desenvolveram a doença.
- a + b = total de indivíduos expostos ao fator de risco.
- c + d = total de indivíduos não expostos ao fator de risco.

O **risco relativo** (RR), como medida de força de associação, é obtido a partir dos estudos de *coorte*, já que seu desenho nos permite calcular a incidência da doença em ambos os grupos. O risco relativo é uma *razão de incidências*, ou seja o quociente entre a incidência da doença nos expostos e a incidência nos não expostos ao suposto fator de risco. Isso é:

$$\text{risco relativo} = \frac{\text{incidência em expostos}}{\text{incidência em não expostos}}$$

Na tabela 2x2, corresponde a:

$$\text{RR} = \frac{I_E}{I_{\bar{E}}} = \frac{a/a+b}{c/c+d}$$

**Risco relativo:** razão entre o risco absoluto de adoecer ou morrer daqueles com a exposição de interesse e o risco absoluto de adoecer ou morrer daqueles sem a exposição de interesse.

Um RR igual à unidade (RR=1) é interpretado como a ausência de associação entre exposição e doença (o risco de adoecer é similar em ambos os grupos); um RR maior que 1 (RR>1) indica maior risco nos expostos (a exposição está associada à doença e é um fator de risco); um RR menor que 1 (RR<1) indica menor risco nos expostos (a exposição está associada à doença e é um fator protetor). A magnitude do RR quantifica a *força de associação* entre a exposição e a doença, assim um RR igual a 3,5 expressa uma associação mais forte entre exposição e doença que, por exemplo, um RR igual a 1,4, ou um RR igual a 0,2 indica uma associação mais forte que um RR igual a 0,7.

Consideremos um exemplo clássico. Entre 1950 e 1952, os doutores Dawber, Meadors e Moore do Serviço de Saúde Pública dos EUA selecionaram 5.127 homens e mulheres saudáveis, de 30 a 59 anos, residentes em Framingham, Massachusetts, que desde então estavam sendo estudados prospectivamente com a finalidade de observar a relação entre uma série de fatores de risco e o desenvolvimento de doença cardíaco-coronária. Como



parte do Estudo do Coração de Framingham, a Tabela 3.16 apresenta a situação observada no décimo sexto ano de segmento de um grupo de 1.112 homens saudáveis, de 35 a 44 anos de idade, no início do estudo, em relação ao desenvolvimento de doenças cardíaco coronárias (DCC) conforme a sua exposição a três fatores de risco selecionados.

**Tabela 3.16** Exposição a fatores de risco e doença cardíaco coronária (DCC). Segmento de 16 anos em homens de 35-44 anos de idade. Framingham, EUA

Coorte	Hipertensão		Cardiomegalia		Tabagismo	
	Total	DCC	Total	DCC	Total	DCC
Expostos	22	12	111	41	800	181
Não-expostos	1.090	206	1.001	177	312	37

Vejam os primeiros fatores de risco estudados: hipertensão (definida aqui como pressão arterial sistólica igual ou maior que 180 mmHg). De acordo com os dados, dos 1.112 homens no início do estudo, 22 tinham hipertensão (estavam expostos) e 1.090 não tinham (não estavam expostos). Dezesesseis anos depois, 12 dos expostos e 206 dos não expostos (logo, 218 casos) tinham desenvolvido doença cardíaco coronária (DCC). Isso quer dizer que o risco absoluto de adoecer com DCC entre os que têm hipertensão é:

$$\text{incidência em expostos: } \frac{12}{22} \times 1.000 = 545,5 \text{ por } 1.000 \text{ expostos}$$

Já o risco absoluto de adoecer com DCC entre os que não têm hipertensão é:

$$\text{incidência em não expostos: } \frac{206}{1.090} \times 1.000 = 189,0 \text{ por } 1.000 \text{ não expostos}$$

Agora poderemos comparar ambos os riscos absolutos e determinar o excesso de risco como uma razão de taxas. Assim, o risco relativo será:

$$\text{risco relativo: } \frac{545,5}{189,0} = 2,89$$

Isso é, os indivíduos hipertensos têm 2,89 vezes o risco de adoecer com DCC que os indivíduos não hipertensos. Dito de outro modo, a exposição ao fator de risco incrementa 1,89 vezes o risco de desenvolver a doença. Assim, o risco relativo indica quanto mais risco tem os expostos *em relação* aos não expostos.

Examinemos o seguinte exemplo. Na última quinzena de maio de 1991, apareceu um intenso surto de cólera em três aldeias da Amazônia (população de 1.761 habitantes), que afetou 125 pessoas e provocou a morte de sete delas. Durante a primeira semana

de junho, realizou-se uma enquete rápida em toda a população sobre a exposição a um conjunto de supostos fatores de risco. A Tabela 3.17 apresenta os resultados em relação ao consumo de certos produtos que, suspeitava-se, poderiam estar envolvidos no surto.

**Tabela 3.17** Surto de cólera em três aldeias rurais, Amazônia; Junho 1991.

Suposto fator de risco	Adoeceram (n=125)		Não adoeceram (n=1.636)	
	Expostos	Não expostos	Expostos	Não expostos
Água não tratada	111	14	1.093	543
Peixe cru	7	118	14	1.622
Peixe cozido	17	108	198	1.438
Arroz requentado	47	78	522	1.114
“Tamal” de arroz*	24	101	272	1.364
Fruta sem lavar	71	54	683	953

\*Tamal - prato típico

Objetivamos determinar se existe alguma *associação* entre a exposição a supostos fatores de risco e a presença de cólera na população. Para isso, procedemos a construir uma tabela 2x2 para *cada um* dos supostos fatores de risco, a saber:

Água não tratada	Doente	Não doente	Total
exposto	111	1.093	1.204
não exposto	14	543	557
Total	125	1.636	1.761

Peixe cru	Doente	Não doente	Total
exposto	7	14	21
não exposto	118	1.622	1.740
Total	125	1.636	1.761

Peixe cozido	Doente	Não doente	Total
exposto	17	198	215
não exposto	108	1.438	1.546
Total	125	1.636	1.761

Arroz requentado	Doente	Não doente	Total
exposto	47	522	569
não exposto	78	1.114	1.192
Total	125	1.636	1.761

“Tamal” de arroz	Doente.	Não doente	Total
------------------	---------	------------	-------

exposto	24	272	296
não exposto	101	1.364	1.465
Total	125	1.636	1.761

Fruta sem lavar	Doente.	Não doente	Total
exposto	71	683	754
não exposto	54	953	1.007
Total	125	1.636	1.761

Agora poderemos calcular com mais facilidade as medidas de associação, assumindo que esse é um estudo de *coorte*. Portanto, os riscos relativos em cada exposição seriam os seguintes:

$$RR_{\text{água não tratada}} = \frac{\frac{111}{14}}{\frac{1.204}{557}} = \frac{0,092}{0,025} = 3,67$$

$$RR_{\text{arroz requentado}} = \frac{\frac{47}{78}}{\frac{569}{1.192}} = \frac{0,083}{0,065} = 1,26$$

$$RR_{\text{peixe cru}} = \frac{\frac{7}{21}}{\frac{118}{1.740}} = \frac{0,333}{0,068} = 4,92$$

$$RR_{\text{tamal de arroz}} = \frac{\frac{24}{296}}{\frac{101}{1.465}} = \frac{0,081}{0,069} = 1,18$$

$$RR_{\text{peixe cozido}} = \frac{\frac{17}{215}}{\frac{108}{1.546}} = \frac{0,079}{0,070} = 1,13$$

$$RR_{\text{fruta sem lavar}} = \frac{\frac{71}{754}}{\frac{54}{1.007}} = \frac{0,094}{0,054} = 1,76$$

Parece evidente que a exposição que está associada com mais força ao fato de adoecer é o consumo de peixe cru, assim como o consumo de água não tratada. O risco relativo de 4,92 nos indica que a probabilidade de desenvolver cólera foi 3,92 vezes maior nos sujeitos que consumiram peixe cru que nos que não consumiram. O risco de apresentar cólera foi também quase três vezes maior naqueles que consumiram água não tratada. Considerando os riscos relativos restantes, essa evidência aponta para uma implicação causal desses produtos no surto de cólera nessas comunidades amazonenses.



### Exercício 3.6

Para quantificar a associação entre a exposição de mulheres gestantes a uma série de supostos fatores de risco e a presença de baixo peso ao nascer (BPN) em seus conceptos, entre maio e setembro de 1996, realizou-se um estudo epidemiológico na Província N de um país da América do Sul. Definiu-se como doente toda a mãe que tinha tido um menino/a com BPN (menos de 2.500 gramas) e como não doente toda a mãe que tinha tido um menino/a sem BPN. Foram selecionados 1.556 doentes e 16.910 não doentes, registrados entre 1988 e 1995 na base de dados do sistema informatizado perinatal disponível em tal província. A Tabela 3.18 apresenta um resumo dos resultados do estudo.

**Tabela 3.18** Prevalência de exposição (%) a fatores de risco de baixo peso ao nascer. Estudo epidemiológico; Província N, América do Sul; 1988-1995.

Fator de risco	Doente (n=1.556)	Não doente (n=16.910)
Mãe adolescente (menor de 17 anos)	13,1	7,1
Doença da placenta	24,4	7,2
Sem acompanhamento pré-natal	15,6	2,1
Atenção pré-natal tardia (a partir do 5º mês)	56,8	31,0

- a) Preencha adequadamente com os dados nas tabelas 2x2 e verifique a possível associação entre exposição e doença.





- b) Interprete e sintetize seus resultados.

## Razão de Chances (*Odds Ratio*)

Como acabamos de ver, para calcular o risco relativo precisamos da incidência da doença nos expostos e não expostos e estas são obtidas através de estudo de *coorte*. Com mais frequência, ocorre que quando precisamos identificar uma associação entre exposição e doença nos encontramos ante uma série de indivíduos que *apresentaram a doença*. Nessa situação, podemos recorrer a um estudo do tipo caso-controle, no qual se compara a *história* de exposição dos doentes com a de um grupo similar de indivíduos, porém sadios, o que se chama “grupo controle”. O desenho do caso-controle é muito versátil e popular no campo e é um dos estudos que nos permite obter uma medida denominada **razão de chances** (razão de produtos cruzados, razão de vantagens, razão de sorte, razão de possibilidades, razão de *odds*), análoga ao risco relativo, que estima o excesso de risco.

Nos estudos caso-controle, parte-se de dois grupos de indivíduos, um com a doença e outro sem ela, e é investigado se tinham estado *previamente* expostos ao fator de risco. Assim, nos estudos de caso-controle, a tabela 2x2 tem os seguintes componentes:

**Tabela 3.19** Tabela 2x2 (Estudo caso-controle)

	Caso	controle	
Exposto	a	b	a + b
Não exposto	c	d	c + d
	a + c	b + d	a + b + c + d

a = doentes (casos) que estiveram expostos ao fator de risco.

b = não doentes (controles) que estiveram expostos ao fator de risco.

c = doentes (casos) que não estiveram expostos ao fator de risco.

d = não doentes (controles) que não estiveram expostos ao fator de risco.

a + c = total de indivíduos doentes (casos).

b + d = total de indivíduos não doentes (controles).

a + b = total de indivíduos expostos ao fator de risco.

c + d = total de indivíduos não expostos ao fator de risco.

No seu original sentido inglês (*odds ratio*), a razão de chances literalmente significa “razão de ‘odds’”. Para compreender melhor o seu significado, devemos revisar, mais uma vez, o termo *probabilidade*. A definição segundo a estatística frequentista nos diz que a probabilidade da ocorrência de um evento indica a frequência relativa limite com que dito evento terá lugar a longo prazo, em testes repetidos em condições similares. O termo “frequência relativa” implica que o valor numérico de qualquer probabilidade está situado entre 0 e 1.

Como vimos ao longo desse Módulo, a probabilidade é muito utilizada para quantificar a frequência esperada de certas características de interesse, sob condições de incerteza como, por exemplo, a porcentagem de indivíduos nos quais se espera que estejam presentes um risco ou uma doença concretos. Essa mesma informação está contida em outra medida, relacionada com a probabilidade, mas que se expressa de forma diferente. Trata-se do *odds*. A *odds* (ou “vantagem”) está definida como probabilidade de que ocorra um evento dividido pela probabilidade de que não ocorra, isso é, o *odds* vem a ser uma razão de probabilidades complementar. Isso é:

$$\text{odds} = \frac{\text{probabilidade do evento}}{1 - \text{probabilidade do evento}}$$

Por exemplo, dizer que nossa equipe de futebol tem uma probabilidade de 80% de ganhar os jogos de domingo é o mesmo que dizer que o seu *odds* é 4 a 1 (80:20) para essa ocasião. Ou, se 60 fumantes desenvolvem tosse crônica e 40 não, o *odds* entre esses 100 fumadores a favor do desenvolvimento da tosse é 60:40, ou 1,5; em contraste, a probabilidade de que esses fumantes desenvolvam tosse é 60/100 ou 0,6.

Em um estudo caso-controle definem-se os *odds* a favor da exposição ao fator de risco. Assim, nos casos, o *odds* de ter estado exposto será:

$$\text{odds}_{\text{casos}} = \frac{a/a + c}{c/a + c} = \frac{a}{c}$$

e, nos controles, será:

$$\text{odds}_{\text{controles}} = \frac{b/b + d}{d/b + d} = \frac{b}{d}$$

Dessa forma, a razão de chance (OR) é o quociente entre os *odds*, em favor da exposição dos casos e dos controles, isso é:

$$\text{OR} = \frac{a/a + c / c/a + c}{b/b + d / d/b + d} = \frac{a/c}{b/d} = \frac{ad}{bc}$$

**Razão de chances:** é o quociente entre dois *odds*. Define-se de forma diferente segundo a situação avaliada. Em um estudo de caso-controle, a razão de chances (OR) é o quociente entre as vantagens (*odds*) a favor da exposição nos casos ( $a/c$ ) e nos controles ( $b/d$ ) (OR de exposição). Em um estudo de *coorte* ou um estudo transversal a *odds* é o quociente entre as vantagens (*odds*) a favor da doença nos expostos ( $a/b$ ) e nos não expostos ( $c/d$ ) (*odds* da doença). Em ambos os casos o cálculo fica reduzido a  $ad/bc$

A razão de chances dos estudos de caso-controle proporciona uma medida que é conceitual e matematicamente análoga ao risco relativo dos estudos de *coorte*. Desde um ponto de vista mais prático, o OR corresponde à razão de produtos cruzados em uma tabela 2x2, como a apresentada nesse Módulo e calcula-se mediante a seguinte fórmula:

$$OR = \frac{a \times d}{b \times c}$$

Essa medida da força da associação tem a mesma interpretação que o risco relativo e em determinadas circunstâncias (de baixa frequência da doença) constitui uma boa aproximação desse. Assim, um OR igual a 1 (OR=1) indica a ausência de associação exposição-doença; um OR maior de 1 (OR >1) indica exposição de risco e um OR menor de 1 (OR<1) indica um efeito protetor.

Retomando o nosso exemplo sobre a cólera na Amazônia, assumamos que a situação corresponde a um desenho do tipo caso-controle, como poderia ser no caso de que se tratasse de consultas em nosso centro de saúde. Em tal situação, os casos seriam todos os doentes com cólera atendidos no centro de saúde e os controles, por exemplo, todos os pacientes atendidos por outra causa. Ainda que assim não possamos calcular a incidência, poderemos sim medir a força de associação entre exposição e doença através da razão de chances (OR), a saber:

$$OR_{\text{água não tratada}} = \frac{111 \times 543}{1.093 \times 14} = \frac{60.273}{15.302} = 3,94 \quad OR_{\text{arroz requentado}} = \frac{47 \times 1.114}{522 \times 78} = \frac{52.358}{40.716} = 1,29$$

$$OR_{\text{peixe cru}} = \frac{7 \times 1.622}{14 \times 118} = \frac{11.354}{1.652} = 6,87 \quad OR_{\text{tamal de arroz}} = \frac{24 \times 1.364}{272 \times 101} = \frac{32.736}{27.472} = 1,19$$

$$OR_{\text{peixe cozido}} = \frac{17 \times 1.438}{198 \times 108} = \frac{24.446}{21.384} = 1,14 \qquad OR_{\text{fruta sem lavar}} = \frac{71 \times 953}{683 \times 54} = \frac{67.663}{36.882} = 1,83$$

Se compararmos a força de associação medida com o RR e com o OR, observaremos que as diferenças não mudam a conclusão sobre as exposições que parecem estar causalmente implicadas na propagação da cólera na comunidade.

É necessário reiterar que tanto o risco relativo como a razão de chances mede o excesso de risco nos expostos com relação aos não expostos a um determinado fator e ambos têm como ponto de referência a unidade (RR=1 ou OR=1). Assim, se, um RR ou um OR é igual a 2, por exemplo, não significa que os expostos tenham duas vezes mais risco que os não expostos, senão uma vez mais, é dizer que os expostos têm o dobro de risco dos não expostos (100% mais risco); um RR ou um OR igual a 1,5 significa que os expostos têm 0,5 vezes mais risco que os não expostos (ou seja, 50% de excesso de risco).



## Apêndice estatístico - epidemiológico

Nesta seção, são apresentados elementos adicionais para a análise epidemiológica, os quais podem ser consultados posteriormente. Esses incluem testes de significância estatística, estimativa de intervalos de confiança, métodos para a análise de impacto de fatores de risco na população e métodos para o controle de fatores que podem distorcer a comparação entre as populações.

### Comparação das Médias: O Teste Z

Depois de um surto de malária, um centro de saúde realiza um programa de triagem ou rastreamento no qual 150 esfregaços sanguíneos de crianças de 1 a 4 anos de idade são examinadas para detectar a presença de parasitos *Plasmodium falciparum*. São encontradas 70 lâminas positivas e o nível médio de hemoglobina nessas crianças foi de 10,6 g/dL, com um desvio-padrão de 1,4 g/dL. O nível médio de hemoglobina nas 80 crianças com lâminas negativas foi de 11,5 g/dL, com um desvio-padrão de 1,3/dL. O centro de saúde estava interessado em saber se a infecção por *P. falciparum* diminuiu os níveis de hemoglobina nas crianças da comunidade. Os dados que resumem a situação são descritos a seguir:

	crianças com malária (grupo 1)	crianças sem malária (grupo 2)
tamanho do grupo (n)	70	80
media ( $\bar{x}$ )	10,6	11,5
desvio-padrão (DP)	1,4	1,3

Do ponto de vista estatístico, o interesse do centro de saúde consiste em saber se existe ou não diferença entre as médias de hemoglobina observadas nos grupos de crianças. Nesse caso, podemos usar um teste de significância estatística chamado Teste Z, cujos valores estatísticos são:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{DP_1^2}{n_1} + \frac{DP_2^2}{n_2}}}$$

Segundo esse teste, o valor crítico do Z para um nível de significância de 0,05 (5%) é 1,96. Quando o valor de Z calculado é maior que o Z crítico (1,96), é rejeitada a hipótese nula e conclui-se que a diferença observada é estatisticamente significativa com 95% de confiança ( $p < 0,05$ ). Substituindo,

$$z = \frac{10,6 - 11,5}{\sqrt{\frac{1,4^2}{70} + \frac{1,3^2}{80}}} \quad z = \frac{-0,9}{0,2216} = -4,06 \quad (p < 0,05)$$

O resultado é significativo em nível 0,05 (5%), posto que 4,06 é maior que 1,96 (nesse teste é considerado o valor absoluto). Consequentemente, pode-se afirmar com 95% de confiança que a infecção por *P. falciparum* diminui os níveis de hemoglobina nas crianças atingidas. Essa confiança é *estatística*, porque expressa a segurança relativa com que se toma uma decisão *estatística*: dizer que há ou não há diferença entre dois grupos observados, em função da influência do acaso. Essa decisão, em estatística, corresponde a rejeitar ou aceitar a *hipótese nula* ou de diferença. A Tabela 3.14 apresenta outros valores críticos de Z para diferentes níveis de significância. Esse teste é usado somente quando ambas as amostras são grandes (>30 cada grupo) de outro modo, costuma-se aplicar o teste *t* de Student, não tratado nesse Módulo.

### Intervalos de Confiança para médias e proporções

Um procedimento alternativo ou complementar da inferência estatística para estimar o grau de incerteza que há entre os estimadores dos parâmetros populacionais é o cálculo de seus **intervalos de confiança**. Sua comparação direta, inclusive, pode ser tanto ou mais informativa que os valores de *p* do nível de significância para decidir se uma diferença entre os estimadores (médias ou proporções) é ou não estatisticamente significativa.

De modo geral, um intervalo de confiança é *simétrico* em relação ao estimador que contém, isso é, que se constrói somando e diminuindo uma mesma quantidade à média ou à proporção observada na população estudada. Essa quantidade se chama *erro de amostragem* e corresponde ao produto do erro-padrão do estimador e um valor crítico do teste estatístico correspondente, que costuma ser o valor de Z para um nível de confiança de 95% (1,96). Assim,

$$\text{intervalo de confiança} = \text{estimador} \pm \underbrace{Z_{95\%} \times \text{erro-padrão (E.P.)}}_{\text{erro de amostragem}}$$

O erro-padrão é o mesmo que se usa nos testes de significância estatística e, como já vimos, é uma expressão da variação ou variabilidade entre os indivíduos nas amostras da população. Mais concretamente, o erro-padrão representa o desvio-padrão de uma distribuição de *amostras* repetidas da mesma população. Imaginemos que em uma mesma população repete-se 100 vezes o mesmo estudo para determinar o período de incubação médio, em dias, de difteria. Cada estudo resumirá os seus resultados em duas medidas:

a média e o desvio-padrão do período de incubação da difteria. Assim, teremos 100 médias e 100 desvios-padrão do mesmo. Se fizermos um gráfico das 100 médias obtidas, veremos que estas seguem uma distribuição normal e, portanto, essa distribuição de amostras terá também uma média e um desvio-padrão. Essa média é uma média de médias e esse desvio-padrão é, justamente, o **erro-padrão**; ambos representam os verdadeiros parâmetros populacionais do período de incubação da difteria. O erro-padrão é uma medida de grande importância na teoria de amostragem; na prática se expressa como uma relação entre a dispersão dos dados observados e o tamanho da amostra estudada.

O intervalo de confiança desse modo construído indica a posição em que o verdadeiro parâmetro populacional estudado está contido, com uma probabilidade conhecida. Por exemplo, se um estudo sobre o período de incubação da difteria reporta que a média é 3,5 dias e o Intervalo de Confiança de 95% (IC95%) é 2,4-6,1 dias, o que estão dizendo os autores é: “apesar do nosso resultado pontual ser de 3,5 dias em média, reconhecemos que essa é uma estimativa e portanto está sujeita a erro aleatório. No entanto, temos 95% de confiança que o verdadeiro tempo de incubação da difteria está entre 2,4 dias e 6,1 dias”. Em outras palavras, estão dizendo que se for repetido 100 vezes o mesmo estudo, em 95 das vezes a média obtida estará entre 2,4 e 6,1 dias, mas em 5 das vezes poderá estar fora do intervalo de confiança. Dessa forma, o intervalo de confiança quantifica de modo explícito a inerente imprecisão dos dados.

### IC95% de uma média:

$$\bar{x} \pm 1,96 \cdot \left( \frac{DP}{\sqrt{n}} \right)$$

O IC95% da média de hemoglobina no grupo de crianças com malária será:

$$10,6 \pm 1,96 \left( \frac{1,4}{\sqrt{70}} \right) = 10,6 \pm 1,96 \times 0,1673 = 10,6 \pm 0,328$$

Limite inferior ( $L_i$ ) = 10,3g/dL; Limite superior ( $L_s$ ) = 10,9g/dL.

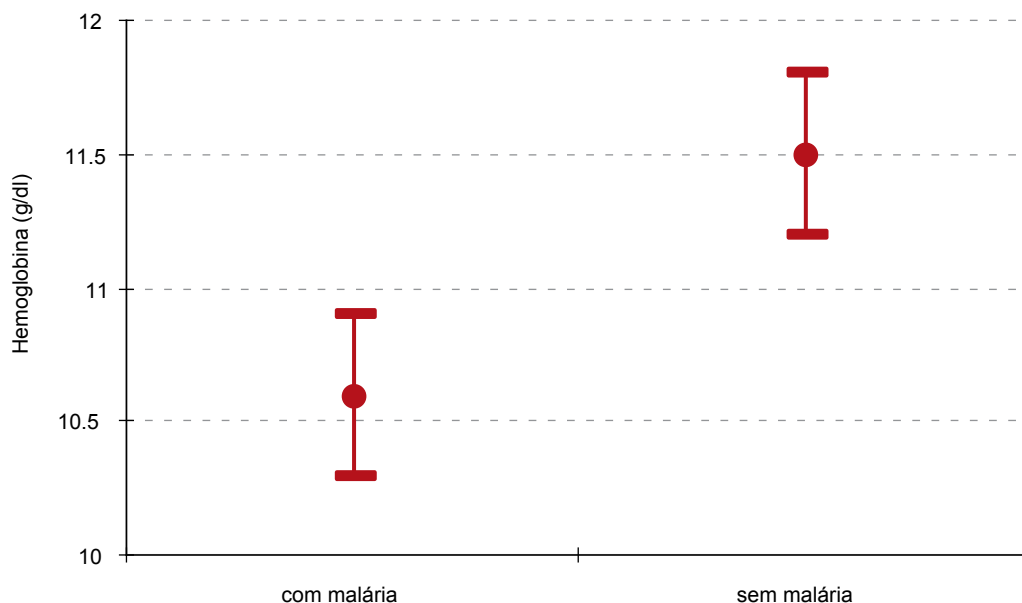
O IC95% da média de hemoglobina no grupo de crianças sem malária será:

$$11,5 \pm 1,96 \cdot \left( \frac{1,3}{\sqrt{80}} \right) = 11,5 \pm 1,96 \times 0,1453 = 11,5 \pm 0,285$$

$L_i$  = 11,2g/dL;  $L_s$  = 11,8g/dL.

O Gráfico 3.10 compara visualmente ambos os grupos de população em função de seus respectivos intervalos de confiança. O fato de observar que suas posições não se sobrepõem entre si (o limite superior de um e o limite inferior de outro “não se tocam”) sugere que há diferença estatisticamente significativa entre ambos, em correspondência com os resultados do teste de significância empregado (Teste Z).

**Gráfico 3.10** Comparação de intervalos de confiança de duas médias



**IC95% de uma proporção:**

$$p \pm 1,96 \sqrt{\frac{p \cdot (1 - p)}{n}}$$

O IC95% de confiança para a incidência de malária em camponeses será:

$$0,0882 \pm 1,96 \sqrt{\frac{0,0882 \cdot (1 - 0,0882)}{204}} = 0,0882 \pm 1,96 \times 0,0199 = 0,0882 \pm 0,0389$$

$L_i = 0,0493$  (49,3 por mil);  $L_s = 0,1272$  (127,2 por mil).

Observe que a proporção  $p$  na fórmula se expressa em um determinado valor (ou seja, vai de 0 a 1); os resultados são multiplicados por 1.000 para serem expressos como taxas por mil.

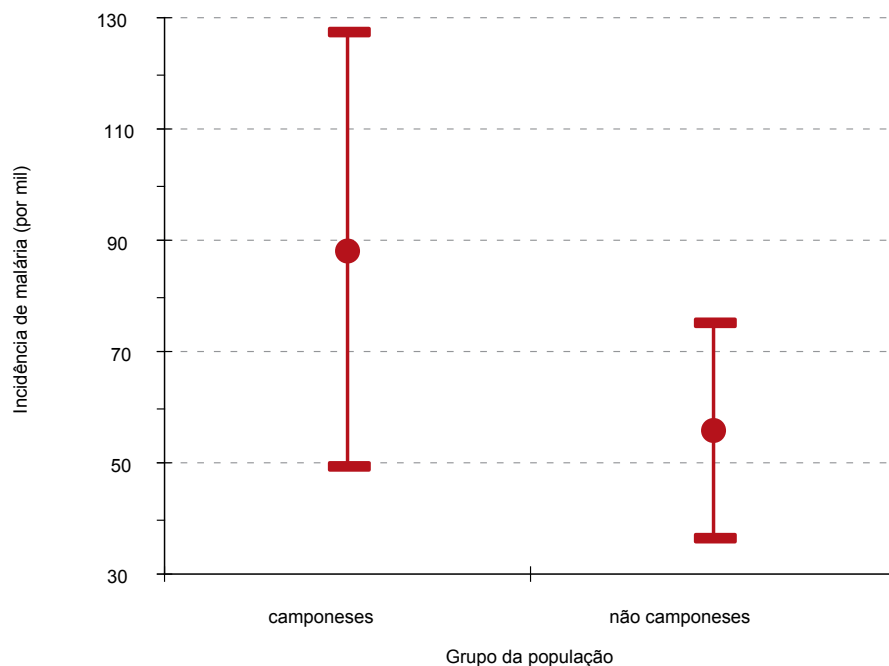
O IC95% de confiança para a incidência de malária entre os não camponeses será:

$$0,0558 \pm 1,96 \sqrt{\frac{0,0558 \cdot (1 - 0,0558)}{556}} = 0,0558 \pm 1,96 \times 0,0097 = 0,0558 \pm 0,0191$$

$L_i = 0,0367$  (36,7 por mil);  $L_s = 0,0748$  (74,8 por mil).

O Gráfico 3.11 compara visualmente ambos grupos populacionais em função de seus respectivos intervalos de confiança. O fato de observar que suas posições se sobrepõem entre si (o limite superior de uma e o limite inferior da outra “se tocam”) sugere que não há diferença estatisticamente significativa entre ambos, em correspondência com os resultados do teste de significância empregado (Chi-Quadrado). Além disso, observe a largura do intervalo de confiança para a incidência em camponeses: a amplitude de um intervalo de confiança também informa sobre a precisão do estudo; os intervalos amplos indicam perda de precisão da amostragem.

**Gráfico 3.11** Comparação de intervalos de confiança de duas proporções



A estimativa de intervalos de confiança é um procedimento básico para documentar o grau de precisão de nossos resultados. Os métodos que revisamos são mais comuns e úteis na prática cotidiana da epidemiologia. No entanto, deve ser ressaltado que existe uma ampla variedade de técnicas para a estimativa de intervalos de confiança de distintos parâmetros populacionais, como os intervalos de confiança de uma diferença de médias, de uma diferença de proporções, riscos relativos, razões de chances e outras importantes medições em saúde. Ainda que o desenvolvimento desses cálculos não faça parte dos objetivos do MOPECE, os princípios básicos que foram descritos sobre a construção de intervalos de confiança são aplicados universalmente para tais casos.

## Medidas de impacto potencial

### Riscos atribuíveis

A comparação de duas medidas de incidência acumulada, como foi destacado, permite quantificar o excesso de risco entre dois grupos com diferentes níveis de exposição ao fator de interesse. Esse excesso de risco pode ser medido em duas direções: a *razão* de taxas, dividindo-as para obter o risco relativo, ou a *diferença* de taxas, diminuindo-as, em cujo caso obteremos o chamado **risco atribuível em expostos**, isso é:

*risco atribuível em expostos = incidência em expostos – incidência em não expostos*

$$RA_E = I_E - I_{\bar{E}}$$

No exemplo de Framingham, o risco atribuível em expostos (hipertensos) será:

risco atribuível em expostos:  $545,5 - 189,0 = 356,5$  por 1.000 expostos

que é a taxa da DCC nos expostos que se considera devida ou atribuível à exposição, isso é, ao fato de ser hipertensos. Em outras palavras, se os indivíduos expostos deixassem de ser hipertensos (isso é, se eliminassem sua exposição ao fator de risco), seu risco absoluto original (545,5 por 1000) ficaria reduzido somente ao risco absoluto dos não expostos (189,0 por 1000); esse excesso de risco (365,5 por 1000) de adoecer com DCC é atribuído à hipertensão.

De forma análoga, o **risco atribuível na população (RAP)** é a magnitude absoluta de incidência de doença que se atribui à exposição, já não somente nos expostos senão *no conjunto da população*. O RAP é a diferença entre a incidência na população ( $I_p$ ) e a incidência nos não expostos, isso é:

Risco atribuível na população (RAP):  $I_p - I_{\bar{E}}$

No exemplo de Framingham, podemos determinar que a incidência de DCC na população foi 196,0 por mil pessoas (218 casos em 1.112 pessoas). O RAP será:

Risco atribuível na população, RAP:  $196,0 - 189,0 = 7,0$  por 1.000 habitantes

Esse valor corresponde à taxa de DCC na população que se considera devida ou atribuível à hipertensão (a exposição). Isso representa que, se não houvesse hipertensão na população, o risco absoluto de adoecer com DCC seria 189,0 por mil pessoas em vez de 196,0 por mil: o excesso de risco, atribuível à exposição, é somente de 7,0 por mil habitantes.

### Frações atribuíveis

A partir dos riscos atribuíveis podemos calcular outras duas medidas importantes, chamadas *medidas de impacto potencial*: a fração atribuível em expostos e a fração atribuível na população. A **fração atribuível em expostos** é simplesmente o risco atribuível em expostos expresso percentualmente, isso é, a proporção da incidência em expostos que se considera devida à exposição ao fator de risco. Isso é:

Fração atribuível em expostos:  $\frac{I_E - I_{\bar{E}}}{I_E} \times 100$

No exemplo de Framingham sobre hipertensão e DCC, a fração atribuível em expostos chamada também risco atribuível percentual, RA%, será:

Fração atribuível em expostos (RA%):  $\frac{545,5 - 189,0}{545,5} \times 100 = 65,4\%$

Isso quer dizer que o 65% do risco absoluto em expostos é devido à exposição ao fator de risco, isso é, ao fato de ser hipertensos (o risco atribuível em expostos, 356,5 por 1.000, equivale a 65,4% do risco absoluto em expostos, 545,5 por 1.000).

Finalmente, a **fração atribuível na população** ou *fração etiológica* é uma medida que quantifica o excesso de risco de adoecer, que se atribui à exposição já não somente nos expostos, senão, *no conjunto da população*. Essa medida é simplesmente o risco atribuível na população expresso percentualmente (RAP%). Dessa forma, a fração atribuível na população é expressa como:

Fração atribuível na população (RAP%):  $\frac{I_P - I_{\bar{E}}}{I_P} \times 100$

A fração atribuível na população permite identificar a importância relativa da exposição a um determinado fator de risco na população, pois expressa a magnitude em que seria reduzido o risco absoluto de adoecer no conjunto da população (isso é, a incidência da doença na população) caso fosse eliminada dita exposição. Por isso, considera-se que o RAP% é uma medida de grande transcendência em saúde pública.

No exemplo de Framingham sobre hipertensão e DCC, a fração etiológica ou risco atribuível populacional percentual RAP%, será:

$$\text{Fração atribuível na população, RAP\%: } \frac{196,0 - 189,0}{196,0} \times 100 = 3,6\%$$

Isso quer dizer que a hipertensão representa somente 3,6% do risco absoluto de adoecer com DCC. Em outras palavras, se tivesse sucesso a eliminação da hipertensão em toda a população, se teria conseguido reduzir unicamente 3,6% o risco de desenvolver doença cardíaco-coronária nessa população.





### Exercício 3.7

Complete a informação do quadro seguinte com relação ao impacto potencial associado aos outros dois fatores de risco de DCC apresentados no Tabela 3.16a:

Fator de risco	Cardiomegalia (medida pelo incremento da sombra cardíaca pelo raio X)	Tabagismo (hábito presente de fumar cigarros)
Medida		
Risco relativo		
Risco atribuível em expostos		
Risco atribuível na população		
Fração atribuível em exposto		
Fração atribuível na população		

Façamos agora um resumo de nossas observações sobre a situação descrita no exemplo de Framingham. Estivemos interessados em ver qual era o risco de desenvolver doença cardíaco-coronária em uma população em função da exposição a três fatores de risco: hipertensão, cardiomegalia e tabagismo. Vimos que o fator com o maior risco relativo é a hipertensão ( $RR=2,9$ ) e o de menor o tabagismo ( $RR=1,9$ ) mas, por outro lado, a fração

etiológica mais alta é a do tabagismo (RAP%=39,5%) e a mais baixa a de hipertensão (RAP%=3,6%). Em outras palavras, ainda que um indivíduo hipertenso claramente tenha mais risco de desenvolver DCC que um fumante, em nível populacional, seria mais recomendável aplicar uma estratégia de redução do tabagismo, do que uma de redução da hipertensão. Por quê?

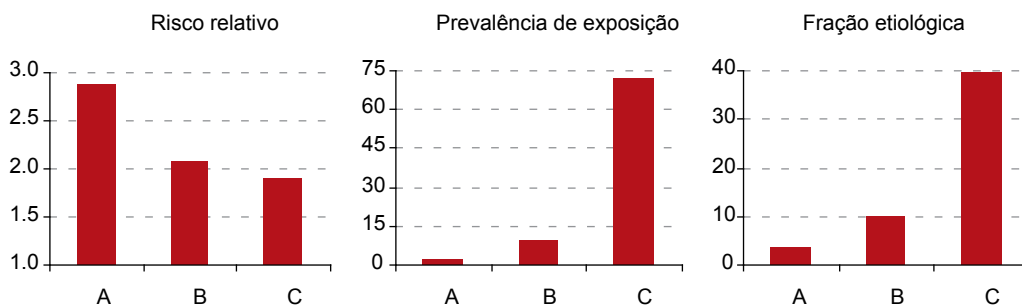
Essa situação, relativamente comum, ilustra o contraste entre risco individual e risco populacional. Essa situação se explica pelas diferenças na **prevalência de exposição na população** aos distintos fatores de risco. Com os dados do Tabela 3.21, podemos inferir que a prevalência de hipertensão na população foi de 2%, enquanto que a de tabagismo foi de 72%. Assim, o RAP% é sensível não somente à magnitude do excesso de risco entre expostos e não expostos, senão também à magnitude da exposição ao risco na população. No exemplo de Framingham, que sintetizamos no Tabela 3.21 e Gráfico 3.12 seguintes, se tivéssemos que decidir por uma intervenção de saúde pública para diminuir o risco de DCC – uma situação habitual quando há limitação de recursos – a intervenção dirigida para diminuir a prevalência do consumo de cigarros terá definitivamente maior impacto em dita população.

**Tabela 3.21** Exposição a fatores de risco e doença cardíaco-coronária (DCC). Segmento de 16 anos em homens de 35-44 anos de idade. Framingham, EUA.

Fator de risco inicial		Risco relativo (RR)	Prevalência de exposição (P <sub>E</sub> )	Fração etiológica (RAP%)
A	hipertensão sistólica	2,9	2,0	3,6
B	cardiomegalia	2,1	10,0	9,8
C	tabagismo	1,9	71,9	39,5

Adaptado de Kahn e Sempos.

**Gráfico 3.12** Exposição a fatores de risco e doença cardíaco-coronária (DCC). Segmento de 16 anos em homens de 35-44 anos de idade. Framingham, EUA.



Outro ilustrativo exemplo da importância de considerar o impacto potencial das medidas de controle na comunidade na hora de tomar decisões encontra-se no caso da

cólera na Amazônia, apresentado no Tabela 3.17. De acordo com a informação disponível, podemos determinar a fração atribuível na população (fração etiológica, RAP%) da exposição a peixe cru e água não tratada:

$$\text{RAP\%}_{\text{peixe cru}}: \frac{\left(\frac{125}{1.761}\right) - \left(\frac{118}{1.740}\right)}{\frac{125}{1.761}} \times 100 = \frac{(0,071 - 0,068)}{0,071} \times 100 = 4,46\%$$

$$\text{RAP\%}_{\text{água não tratada}}: \frac{\left(\frac{125}{1.761}\right) - \left(\frac{14}{557}\right)}{\frac{125}{1.761}} \times 100 = \frac{(0,071 - 0,025)}{0,071} \times 100 = 64,59\%$$

Isso nos permite saber que, apesar do consumo de peixe cru estar mais fortemente associado à presença de cólera (RR=4,9), seu impacto potencial na população é muito baixo, porque a prevalência de exposição (consumo de peixe cru) na população é muito baixa (1,2%). No caso de eliminar o consumo de peixe cru *nessa* população, somente conseguiríamos reduzir aproximadamente 5% de sua incidência de cólera. Pelo contrário, o consumo de água não tratada - uma prática altamente prevalente em dita comunidade (68,4% consome água não tratada) - tem grande impacto potencial; eliminar esse fator de risco reduziria em 65% a incidência de cólera na população. Ainda que seja óbvio que as campanhas de prevenção de cólera nessa comunidade terão de comunicar o risco de consumir peixe cru ou frutas sem lavar, o investimento terá que se dirigir prioritariamente para evitar que sua população consuma água não tratada.

### Padronização de taxas

Com a finalidade de sintetizar a informação disponível, frequentemente as taxas são apresentadas para a população completa ou para grandes categorias da mesma. Essas taxas são chamadas **taxas brutas**. Com estas taxas-resumo, costuma-se comparar a mortalidade ou a incidência de uma doença entre duas áreas geográficas, dois grupos populacionais ou dois momentos no tempo e ver se existem diferenças relevantes. Consideremos a seguinte situação:

Faixa etária (anos)	PAÍS A			PAÍS B		
	Óbitos	População	Taxa por mil	Óbitos	População	Taxa por mil
Total	1.269.166	68.386.000	18,6	5.564.944	198.250.000	28,1

Os dados de mortalidade e população correspondem ao mesmo ano-calendário. O País A é um país pouco industrializado; o País B é um país altamente industrializado. Que informação comparativa pode-se obter a partir desses dados? Basicamente, poderíamos tirar duas conclusões: 1) que a taxa de mortalidade no país B é mais alta que no País A, e 2) que o risco de morrer no País B é 50% mais alto que no País A. Isso significa que, em

termos de mortalidade, o País B está em uma situação mais desfavorável que o País A. O País B, como foi mencionado, é um país altamente industrializado.

As taxas podem também ser apresentadas de forma desagregada para várias categorias da população, definidas sobre a base de características relevantes para uma comparação, tais como idade, sexo, etnia, ocupação ou nível de exposição a determinado fator de risco. Essas taxas são denominadas **taxas específicas**. Calculando que o risco de morrer ou de contrair a grande maioria de doenças está relacionado, geralmente, com a idade e frequentemente difere entre os sexos, a análise da mortalidade ou a incidência de doença em uma população deve necessariamente fazer uso das correspondentes taxas específicas. Voltemos agora à situação entre os países A e B.

Faixa etária (anos)	PAÍS A			PAÍS B		
	Óbitos	População	Taxa por mil	Óbitos	População	Taxa por mil
Total	1.269.166	68.386.000	18,6	5.564.944	198.250.000	28,1
<15	317.308	19.831.740	16,0	94.169	24.781.250	3,8
15 – 44	338.100	35.218.790	9,6	380.430	79.256.250	4,8
45 – 64	270.261	10.941.760	24,7	1.223.875	61.501.250	19,9
65 e +	343.497	2.393.710	143,5	3.866.470	32.711.250	118,2

A partir dessa informação é possível identificar pelo menos três fatos relevantes: 1) o País A tem taxas de mortalidade específicas por idade mais altas que o País B em *todos* os grupos de idade considerados; 2) a estrutura por idades difere de forma marcante entre as duas populações: o país A concentra sua população em idades jovens; o país B em idades mais velhas; e 3) há uma aparente contradição entre o que essa informação mostra e o que se conclui observando as taxas brutas de mortalidade dos dois países.

Como pode ser explicada essa aparente *confusão*? Dado que, como foi mencionado, o risco de morrer ou adoecer está normalmente associado com a idade, as taxas brutas de mortalidade e incidência dependem criticamente da composição etária de uma população. Isso ganha maior relevância quando o objetivo é comparar duas populações. Pode ser, portanto, impróprio usar taxas brutas para comparar duas populações *distintas* a menos que tenham a mesma estrutura por idades. A diferença de composição etária (ou seja, a variável idade) exerce um *efeito de confusão* na comparação de taxas brutas de mortalidade por país. De fato, uma taxa crua é basicamente uma média ponderada das taxas específicas por categoria, sendo que os pesos são as proporções de população em cada categoria.

Para poder realizar uma comparação livre de distorções que podem produzir diferenças na composição da população, devem ser empregadas as **taxas padronizadas**. As taxas padronizadas ou ajustadas são taxas-resumo construídas estatisticamente para conside-

rar e remover o potencial efeito de confusão da variável idade ou outra terceira variável, ao comparar as taxas de mortalidade ou incidência de duas populações diferentes.

O procedimento básico para o ajuste de taxas (o chamado método direto) precisa contar com as taxas específicas por categoria da variável a ajustar (por exemplo, a idade) em ambas as populações a comparar. São necessários dois conjuntos de taxas específicas por idade (um para cada país) e uma população padrão. A idéia geral é ver qual seria o número total de óbitos que haveria em cada um dos dois países se esses tivessem exatamente a mesma estrutura de idade (a população padrão) e ocorressem as taxas específicas por idade realmente *observadas* de cada país. Em outras palavras, se trata de ajustar a estrutura de mortalidade observada em cada país a uma estrutura de idade única e padrão para os dois. O procedimento inclui os quatro passos seguintes:

1. Fixar a população padrão: pode-se selecionar uma já conhecida ou pode-se construir uma a partir dos dados; por exemplo, somando as populações específicas por categoria de idade em cada grupo;

Faixa etária	PAÍS A	PAÍS B	População padrão (A + B)
Total	68.386.000	198.250.000	266.636.000
< 15	19.831.740	24.781.250	44.612.990
15 – 44	35.218.790	79.256.250	114.475.040
45 – 64	10.941.760	61.501.250	72.443.010
65 e +	2.393.710	32.711.250	35.104.960

2. Calcular o número de óbitos em cada categoria de idade da população-padrão aplicando as taxas específicas por idade observadas em cada uma das duas populações a comparar;

Faixa etária	População padrão	PAÍS A		PAÍS B	
		Taxa observada	Óbitos esperados	Taxa observada	Óbitos esperados
Total	266.636.000				
<15	44.612.990	16,0	713.808	3,8	169.530
15 – 44	114.475.040	9,6	1.098.959	4,8	549.480
45 – 64	72.443.010	24,7	1.789.339	19,9	1.441.616
65 e +	35.104.960	143,5	5.037.556	118,2	4.149.407

Por exemplo, o número de óbitos esperados no grupo de menores de 15 anos para o País A é calculado através de uma regra de três simples: se ocorrem 16 óbitos por cada mil pessoas, quantos óbitos ocorrerão em 44.612.990 pessoas? Isso é,

número de óbitos esperados em <15 anos, País A:  $\frac{44.612.990}{1.000} \times 16 = 713.808$

3. Obter o número total de óbitos esperados em cada país, somando os resultados correspondentes do passo anterior, e,

Faixa etária	População padrão	PAÍS A		PAÍS B	
		Taxa observada	Óbitos esperados	Taxa observada	Óbitos esperados
Total	266.636.000		<b>8.639.663</b>		<b>6.310.033</b>
<15	44.612.990	16,0	713.808	3,8	169.530
15 – 44	114.475.040	9,6	1.098.959	4,8	549.480
45 – 64	72.443.010	24,7	1.789.339	19,9	1.441.616
65 e +	35.104.960	143,5	5.037.556	118,2	4.149.407

4. Calcular as respectivas taxas ajustadas por idade para cada população, dividindo o número total de casos esperados obtidos no passo anterior, entre o total da população padrão.

taxa ajustada de mortalidade, País A:  $\frac{8.639.663}{266.636.000} \times 1.000 = 32,4$  por mil

taxa ajustada de mortalidade, País B:  $\frac{6.310.033}{266.636.000} \times 1.000 = 23,7$  por mil

Comparemos mais uma vez os resultados brutos e ajustados:

Taxa de mortalidade por mil	PAÍS A	PAÍS B
Bruta	18,6	28,1
Padronizada	32,4	23,7

Após ter removido a distorção produzida pela diferença na estrutura de idades, conta-se com uma medida-resumo válida para comparar a mortalidade entre os dois países: a taxa de mortalidade é aproximadamente 40% mais alta no País A que no País B. Cabe ressaltar que o uso de taxas padronizadas somente está indicado com fins comparativos: sua construção estatística está baseada na seleção arbitrária de um padrão e, por isso, a magnitude da cifra necessita do valor intrínseco. Por último, a padronização de taxas não supre as deficiências na qualidade, na cobertura, nem no registro dos dados.

## Referências bibliográficas

- Alleyne GAO. Gente sana en entornos saludables. Informe Anual del Director OPAS 1998. Organización Panamericana de la Salud; Washington DC, 1999.
- Alleyne GAO. Emerging diseases—What now?. *Emerging Infectious Diseases* 1998 July; 4(3):498-500.
- Beaglehole R, Bonita R, Kjellström T. *Epidemiología básica*. Organización Panamericana de la Salud; Washington DC, 1994.
- Bortman M. Factores de riesgo de bajo peso al nacer. *Revista Panamericana de Salud Pública/Pan American Journal of Public Health* 1998 May;3(5):314-21.
- Castillo-Salgado C [Editor]. *Manual sobre el enfoque de riesgo en la atención maternoinfantil*. Segunda edición. Organización Panamericana de la Salud; Washington DC, 1999.
- Centers for Disease Control. *Epidemiologic notes and reports: Eosinophilia-Myalgia Syndrome—New Mexico. Mortality and Morbidity Weekly Report* 1989 November 17,38(45):765-67.
- Colton T. *Estadística en medicina*. Ediciones Díaz de Santos S.A.; Madrid, 1995.
- Donaldson RJ, Donaldson LJ. *Medicina comunitaria*. Editorial Díaz de Santos; Madrid, 1989.
- Fletcher RH, Fletcher SW, Wagner EH. *Epidemiología clínica: aspectos fundamentales*. Segunda edición. Masson, Williams & Wilkins España, S.A.; Madrid, 1998.
- Freund JE, Williams FJ. *Dictionary/Outline of basic statistics*. Dover Publications, Inc.; New York, 1991.
- Gordis L. *Epidemiology*. Second edition. W.B. Saunders Company; Philadelphia, 2000.
- Gore SM, Altman DG. *Statistics in practice*. British Medical Association; London, 1982.
- Kahn HA, Sempos CT. *Statistical methods in epidemiology*. Oxford University Press; New York, 1989.
- Kirkwood BR. *Essentials of medical statistics*. Backwell Scientific Publications; Osney Mead, 1988.
- Last J [Editor]. *A dictionary of epidemiology*. Third edition. International Epidemiological Association. Oxford University Press; New York, 1995.
- Lwanga SK, Tye CY, Ayeni O. *Teaching health statistics. Lesson and seminar outlines*. Second edition. World Health Organization; Geneva, 1999.
- Newcomer RJ, Benjamin AE. *Indicators of chronic health conditions. Monitoring community-level delivery systems*. The Johns Hopkins University Press; Baltimore, 1997.
- Norell SE. *Workbook of epidemiology*. Oxford University Press, Inc.; New York, 1995.
- Norman GR, Streiner DL. *Bioestadística*. Mosby/Doyma Libros, S.A.; Madrid, 1996.
- Organización Panamericana de la Salud. *Las condiciones de salud en las Américas. Volumen I*. Edición de 1994. Organización Panamericana de la Salud; Washington DC, 1994.

Quick RE, Vargas R, Moreno D, et al. Epidemic cholera in the Amazon: the challenge of preventing deaths. *American Journal of Tropical Medicine and Hygiene* 1993;48(5):597-602.

Rosner B. *Fundamentals of Biostatistics*. Fifth Edition. Duxbury Thompson Learning; Pacific Grove, 2000.

Silva LC. *Cultura estadística e investigación científica en el campo de la salud: una mirada crítica*. Ediciones Díaz de Santos, S. A.; Madrid, 1997.

Tukey JW. *Exploratory Data Analysis*. Addison Wesley Publishing Co. 1977.

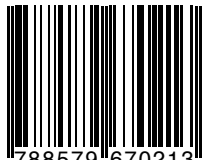
Turnock BJ. Good news and bad news for public health. *Journal of Public Health Management and Practice* 1997 May;3(3):x-xi.

Wallgren A, Wallgren B, Persson R, Jorner U, Haalad JA. *Graphing statistics & data. Creating better charts* SAGE Publications, Inc; Newbury Park, 1996.





ISBN: 978-85-7967-021-3



9 788579 670213



**Organização  
Pan-Americana  
da Saúde**

*Escritório Regional para as Américas da  
Organização Mundial da Saúde*

**Ministério  
da Saúde**

