

Tipos de dados e formas de apresentação na pesquisa clínico-epidemiológica

Types of data and data presentation in clinical-epidemiological research

JOÃO LUIZ DORNELLES BASTOS¹
RODRIGO PEREIRA DUQUIA^{1,2}

Este texto objetiva familiarizar o leitor com os tipos de dados epidemiológicos, assim como suas formas de apresentação tabular e gráfica, com vistas a facilitar a compreensão de trabalhos científicos e a construção de resultados de estudos clínico-epidemiológicos.

De forma a cumprir os propósitos mencionados, o presente artigo está dividido em quatro partes elementares. Na primeira delas, os principais tipos de dados são descritos e exemplificados. A segunda parte, onde são colocadas algumas formas de apresentação de dados em gráficos e em tabelas, é seguida pela listagem de normas básicas sobre como sintetizar e apresentar dados de estudos científicos. Encerra-se a nota com uma breve discussão acerca da importância destes conhecimentos para a condução de uma análise estatística adequada, apresentação de um trabalho científico e leitura crítica de artigos em revistas especializadas.

1 TIPOS DE DADOS

Antes de iniciar propriamente esta seção, cabe esclarecer alguns conceitos fundamentais para a compreensão desta e das futuras notas, especificamente

no que diz respeito às expressões *dados*, *variáveis*, *constantes* e *observações*.

Dados são aqui entendidos como todas as informações coletadas por um pesquisador durante determinada etapa de um estudo.¹ Dados que variam de um indivíduo para o outro ou num mesmo indivíduo em momentos diferentes no tempo são denominados variáveis.¹ Dito de outra forma, variáveis são características ou atributos passíveis de mensuração, que podem tomar diferentes valores, tais como o sexo, a pressão arterial, a cor dos olhos e a idade dos indivíduos investigados.² Tal expressão é utilizada em contraposição ao termo constante, tendo em vista que este último denota qualquer característica conhecida e invariável nas condições em que está sendo analisada. Por sua vez, o termo observação pode ser utilizado quando se trata de qualquer mensuração realizada em um ou mais indivíduos, tendo-se como referência uma ou mais variáveis. Por exemplo, se estamos trabalhando somente com a variável sexo em uma amostra de 20 adultos e conhecemos a quantidade exata de homens e mulheres existentes no grupo, dizemos que esta variável possui vinte observações. Caso o sexo de algum dos indivíduos não tenha sido

¹ Odontólogo. Mestre em Epidemiologia pela Universidade Federal de Pelotas.

² Dermatologista. Mestre em Epidemiologia pela Universidade Federal de Pelotas.

registrado durante a coleta de dados por algum motivo específico, caracterizando o que se chama de uma observação faltante ou, em inglês, *missing*, a variável em questão apresentará um número menor de observações (necessariamente menor do que 20).

Em se tratando dos tipos de variáveis, assunto principal desta seção, é preciso dizer que existem dois grandes grupos: (a) o grupo das variáveis categóricas ou qualitativas, que se subdividem em dicotômicas, nominais e ordinais; e (b) o grupo das variáveis numéricas ou quantitativas, subdivididas em contínuas e discretas.^{1,3-5}

1.1 Variáveis categóricas

a) *Dicotômicas* ou também denominadas *binárias*: são aquelas com apenas duas categorias, ou seja, com apenas duas possibilidades de resposta. Como exemplo, citam-se as variáveis gravidez (sim ou não), relato de dor nas últimas quatro semanas (sim ou não) e sexo (masculino ou feminino).

b) *Nominais*: apresentam três ou mais categorias destituídas de uma hierarquia entre elas. Por exemplo: país de origem (Brasil, Argentina, Paraguai, entre outros), grupo sanguíneo (A, B, AB e O) e estado conjugal (casado, solteiro divorciado e viúvo).

c) *Ordinais*: contando com três ou mais categorias em que há uma ordem explícita entre elas, tais como classe social (classes A, B, C, D e E), sistemas de classificação e estadiamento de cânceres (I, II, III e IV) e tabagismo (não-fumante, ex-fumante, fumante leve e fumante pesado).

1.2 Variáveis numéricas

a) *Discretas*: quando as observações em questão podem tomar apenas números inteiros. Praticamente todos os exemplos de variáveis discretas constituem contagens de eventos, expressas através dos algarismos 1, 2, 3 e assim por diante. Número de filhos, número de consultas com o médico no último ano e número de batimentos cardíacos são exemplos deste tipo de variável.

b) *Contínuas*: geralmente obtidas por meio de alguma forma de mensuração. São medidas em uma escala contínua, isto é, apresentam tantas casas decimais quantas forem passíveis de registro pelo instrumento de medida utilizado. Exemplo: pressão sangüínea, peso ao nascer, altura e colesterol sérico.

Na Figura 1, apresenta-se um diagrama que visa facilitar o entendimento, a identificação e a classificação das variáveis destacadas acima.

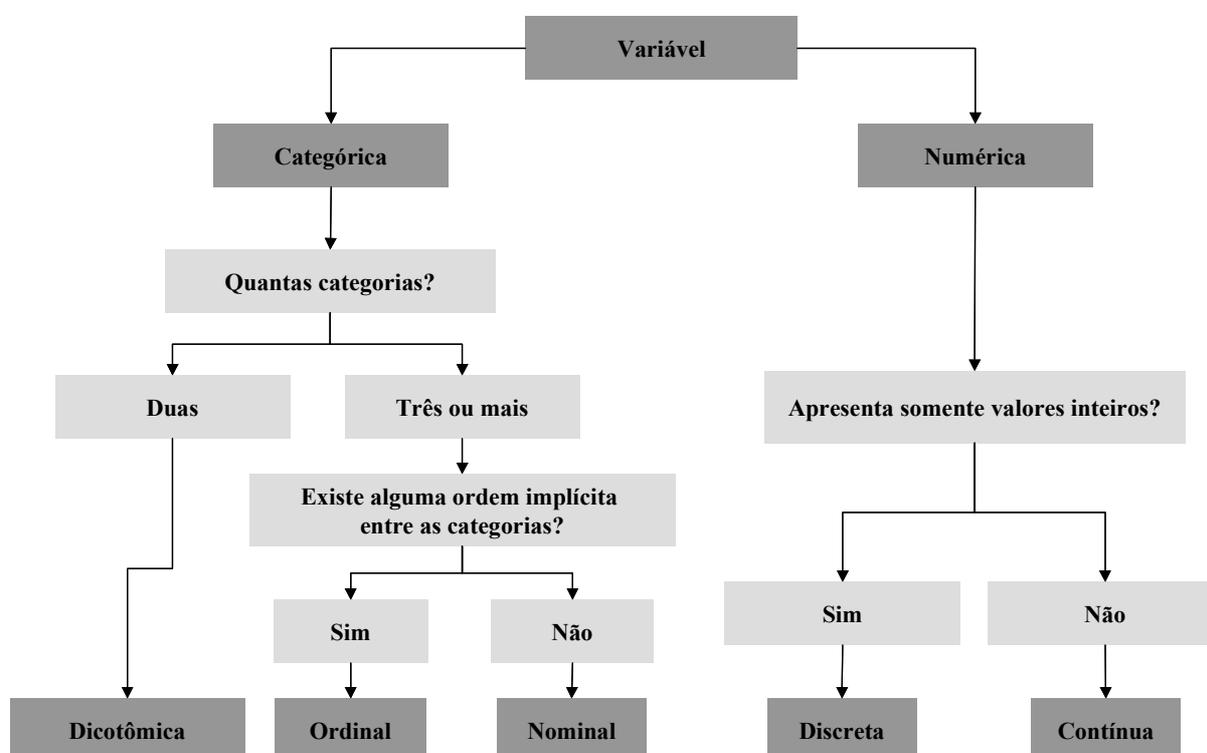


Figura 1 - Diagrama para classificação dos principais tipos de variáveis epidemiológicas.

2 APRESENTAÇÃO GRÁFICA E TABULAR DE DADOS

2.1 Variáveis categóricas

Um dos primeiros passos para analisarmos o comportamento de uma variável consiste na construção de uma distribuição de freqüências, isto é, na organização dos dados conforme a ocorrência dos diferentes resultados em cada categoria.⁴ Para variáveis categóricas, a distribuição de freqüências pode ser apresentada sob a forma de tabelas ou de gráficos, entre eles o gráfico de barras e o de setores ou de *pizza*.⁵

Sintetizar as informações contidas em uma variável categórica por meio de uma tabela consiste em tarefa relativamente fácil. Para tanto, basta contar o número de observações em cada categoria da variável, obtendo-se, assim, suas freqüências absolutas. É bastante útil apresentar as freqüências absolutas acompanhadas de seus respectivos valores percentuais, também chamados de freqüências relativas. Por exemplo, a Tabela 1 expressa em valores absolutos e relativos a ocorrência de dor dentária nos últimos seis meses em um estudo realizado com indivíduos adultos residentes na cidade de Pelotas.⁶

As mesmas informações contidas na Tabela 1 poderiam ser apresentadas através do gráfico de barras ou do gráfico de setores, conforme citado acima. Estes podem ser confeccionados levando-se em conta a freqüência absoluta ou relativa das categorias, cabendo salientar que a forma dos gráficos (comprimento das barras ou tamanho dos setores) não é afetada pelo tipo de freqüência que está sendo considerada, se absoluta ou relativa.⁵ O que deve ser modificado apenas são os rótulos das escalas, que serão diferentes em cada situação (percentual ou quantidade de observações em cada categoria, segundo a escolha). As Figuras 2 e 3 ilustram as informações apresentadas na Tabela 1 sob a forma de um gráfico de barras e de um gráfico de setores, respectivamente.

Tabela 1 - Dor dentária nos últimos seis meses em indivíduos adultos (≥ 20 anos, $n = 2572$). Pelotas, Brasil, 2005.

Dor dentária nos últimos seis meses	Freqüência absoluta (n)	Freqüência relativa (%)
Sim	486	17,66
Não	2266	82,34
Total	2752	100,00

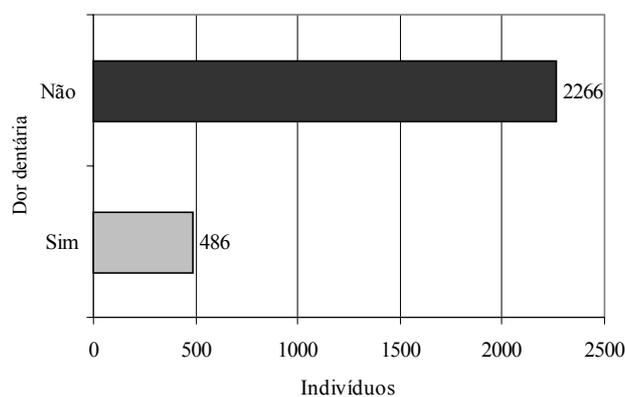


Figura 2 - Dor dentária nos últimos seis meses em indivíduos adultos (≥ 20 anos; $n = 2752$). Pelotas, Brasil, 2005.

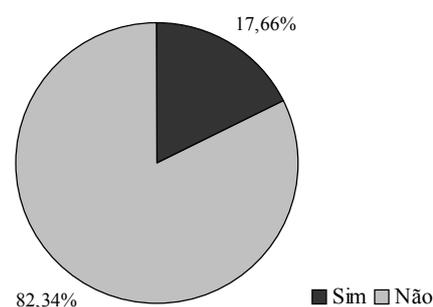


Figura 3 - Dor dentária nos últimos seis meses em indivíduos adultos (≥ 20 anos; $n = 2752$). Pelotas, Brasil, 2005.

2.2 Variáveis numéricas

A distribuição de freqüências de variáveis numéricas pode ser visualizada em uma tabela ou em um gráfico do tipo histograma ou do tipo polígono de freqüências. Para variáveis discretas, é possível apresentar o número de observações de acordo com os diferentes valores encontrados, tal como ilustrado na Tabela 2. Esta tabela apresenta a distribuição de dentes irrompidos na cavidade bucal de crianças com seis meses de idade de um estudo coorte^a iniciado em Pelotas no ano de 1993.⁷ Perceba que, além das freqüências absoluta (n) e relativa (%), estão sendo mostradas as freqüências acumuladas absoluta e relativa. Estas correspondem ao número e ao percentual de indivíduos com valores menores ou iguais a uma contagem específica da variável em questão, respectivamente.¹ Por exemplo, na Tabela 2 há 288 ou 80,9% de crianças com um ou menos dentes irrompidos, 348 ou 97,8% de crianças com dois ou menos dentes irrompidos e assim por diante. Observa-se que a apresentação das freqüências acumuladas é somente possível para variáveis categóricas ordinais ou variáveis numéricas, tal como na tabela apresentada.

TABELA 2 – Número de dentes irrompidos em crianças (n = 359) com seis meses de idade. Pelotas, Brasil, 1993.

Dentes irrompidos	Frequência absoluta (n)	Frequência relativa (%)	Frequência absoluta acumulada	Frequência relativa acumulada
0	273	76,69	273	76,69
1	15	4,21	288	80,90
2	60	16,85	348	97,75
3	1	0,28	349	98,03
4	6	1,69	355	99,72
6	1	0,28	356	100,0
Total	356*	100,00		

* Três crianças não puderam ter seus dentes examinados.

Quando se deseja visualizar a distribuição de frequências de variáveis contínuas, há necessidade de criar intervalos de classe (ou categorias) que agrupem um determinado número de observações em cada um deles.¹ Assim, procede-se contando o número de observações da variável, identificando-se o valor máximo e mínimo e formando-se entre 5 a 20 categorias, preferencialmente de mesmo tamanho, para a apresentação da variável sob a forma tabular ou gráfica.⁵ A Tabela 3 ilustra os valores observados de peso ao nascer (variável numérica contínua, coletada em gramas) em 359 crianças.⁷ e sua distribuição de frequências com os intervalos de classe criados de 1000 em 1000 gramas. A mesma variável em classes de 500 gramas está apresentada em um histograma na Figura 4. O polígono de frequências para a variável peso ao nascer também poderia ser confeccionado. Tendo-se como referência o histograma criado, o polígono de frequências poderia ser desenhado unindo-se os pontos médios das extremidades superiores das barras do histograma por meio de uma linha. Para concluirmos este gráfico, basta unir os extremos da figura formada com o eixo horizontal do gráfico no ponto médio da classe imediatamente inferior e superior à primeira e últi-

TABELA 3 – Distribuição do peso ao nascer de crianças (n = 359) pertencentes a um estudo de coorte. Pelotas, Brasil, 1993.

Peso ao nascer em gramas	Frequência absoluta (n)	Frequência relativa (%)
1000 a 2000	24	6,69
2001 a 3000	148	41,23
3001 a 4000	172	47,90
4001 a 5000	15	4,18
Total	359	100,00

ma classes existentes, respectivamente¹ (vide Figura 5 para um exemplo de polígono de frequências).

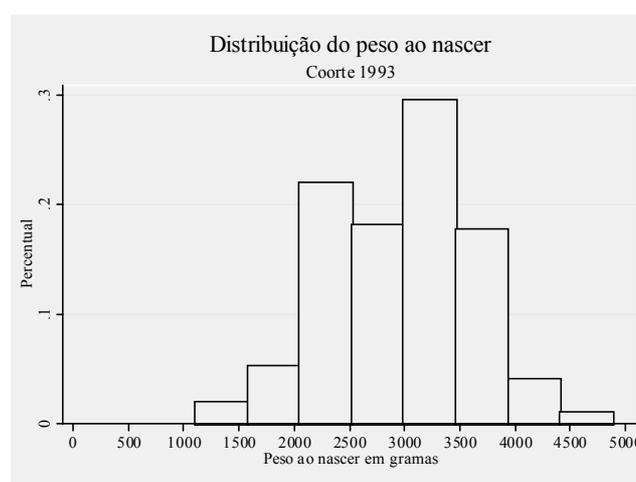


Figura 4 – Distribuição do peso ao nascer em crianças com seis meses de idade (n = 359). Pelotas, Brasil, 1993.

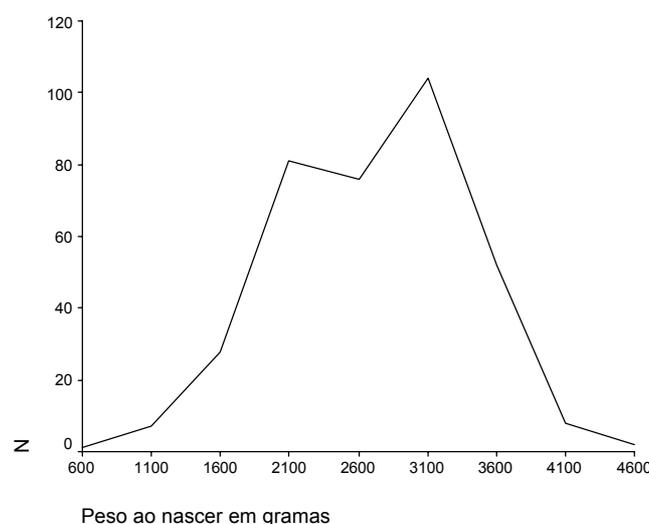


Figura 5 – Distribuição do peso ao nascer em crianças com seis meses de idade (n = 359). Pelotas, Brasil, 1993.

2.3 Tabulações cruzadas para explorar a relação entre duas variáveis

Quando as duas variáveis de interesse são categóricas, pode-se observar a relação entre elas através de uma tabela denominada tabela de contingência. Uma convenção útil é fazer com que as linhas correspondam à variável *exposição* e as colunas à variável *desfecho*.⁵ Isto significa dizer que, se estamos estudando o efeito do tabagismo (exposição) sobre a ocorrência de dor dentária (desfecho) nos últimos seis meses, por exemplo, devemos colocar a variável tabagismo nas linhas e a variável dor dentária nas colunas de uma tabela de contingência. A interpretação desta tabela pode ser facilitada pela inclusão dos totais nas linhas e colunas e dos valores relativos. Os totais devem obedecer à somatória das linhas e/ou das colunas conforme desejado, enquanto que os valores relativos devem estar de acordo com a exposição, isto é, devem totalizar em 100% nas linhas. Assim, a interpretação da Tabela 4 demonstra que na amostra estudada, 25,4%, 13,2% e 15,4% dos fumantes, ex-fumantes e não-fumantes,

respectivamente, relataram dor dentária nos últimos seis meses.⁶ Por outro lado, também se pode afirmar que 74,6%, 86,8% e 84,6% dos fumantes, ex-fumantes e não-fumantes não tiveram dor de dente nos últimos seis meses.⁶ Esta forma de apresentação é amplamente utilizada na literatura e facilita sobremaneira a leitura das tabelas.

Para verificar a relação entre duas variáveis numéricas ou entre uma variável numérica e outra categórica, pode-se utilizar o diagrama de pontos, também chamado de diagrama de dispersão. Neste diagrama, cada par de valores é representado por um símbolo ou um ponto, em que sua posição horizontal é determinada pelo valor da primeira variável e a posição vertical é determinada pelo valor da segunda.⁵ Por convenção, os eixos vertical e horizontal devem corresponder ao desfecho e à exposição, respectivamente (Figura 6, para a relação entre peso e altura em indivíduos adultos residentes na cidade de Pelotas no ano de 2005). Interpreta-se o diagrama da Figura 6 da seguinte forma: o aumento na altura dos indivíduos está acompanhado por um determinado acréscimo em seu peso.

TABELA 4 - Tabagismo e dor dentária nos últimos seis meses em indivíduos adultos (≥ 20 anos; $n = 2752$). Pelotas, Brasil, 2005.

Tabagismo	Dor dentária				Total	
	Sim		Não		N	%
	N	%	N	%		
Fumante	189	25,44	554	74,56	743	100,00
Ex-fumante	75	13,20	493	86,80	568	100,00
Não-fumante	222	15,41	1219	84,59	1441	100,00
Total	486	17,66	2266	82,34	2752	100,00

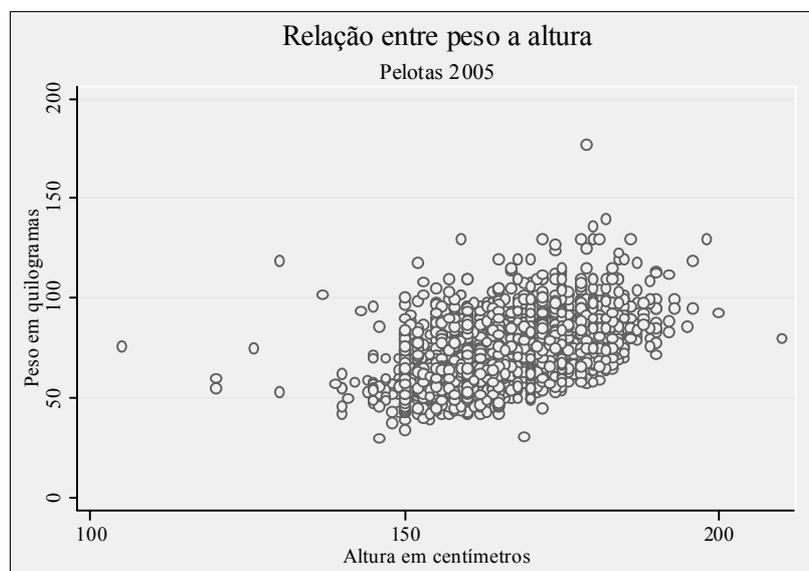


Figura 6 - Diagrama de pontos para a relação entre peso e altura de indivíduos adultos ($n = 3136$) residentes na cidade de Pelotas. Pelotas, RS, Brasil, 2005.

3 NORMAS BÁSICAS PARA A CONFECÇÃO DE TABELAS E DE GRÁFICOS

Idealmente, toda a tabela deve:^{1,4}

- Ser auto-explicativa;
- Exibir valores com o mesmo número de casas de decimais, quando for o caso;
- Conter título informando o que está sendo apresentado, onde e quando foram coletados os dados;
- Possuir estrutura formada por três linhas horizontais, onde duas delimitam o cabeçalho e uma faz o fechamento em sua parte inferior;
- Ser aberta, ou seja, não possuir linhas verticais em suas extremidades laterais;
- Apresentar alguma explicação adicional no rodapé, quando necessário;
- Ser inserida em um documento somente após ter sido mencionada no texto; e
- Ser numerada por algarismos arábicos.

Da mesma forma que as tabelas, os gráficos devem^(1,4):

- Conter título abaixo da figura com todas as informações pertinentes;
- Ser referenciados no texto como figuras;
- Ter seus eixos identificados pelas variáveis analisadas;
- Exibir no rodapé a fonte que forneceu os dados, caso necessário;
- Deixar claro a escala que está sendo utilizada; e
- Ser interpretados sem necessidade de recorrer ao texto.

Além disso, o eixo vertical dos gráficos deve sempre iniciar no valor zero, a fim de evitar possíveis dificuldades de interpretação. Devem também ser evitadas quebras na escala dos eixos e a utilização de efeitos tridimensionais ou cores. Gráficos e tabelas em tons de cinza são suficientes na maioria dos casos.

4 CONSIDERAÇÕES FINAIS

Compreender como se classificam os tipos de dados e como sintetizá-los em tabelas ou gráficos constitui etapa fundamental na observação de aspectos relevantes dos resultados e na elaboração de hipóteses acerca do fenômeno em estudo.⁴ Além disso, esta compreensão é importante durante a etapa de análise do estudo, pois os testes estatísticos disponíveis estão intimamente relacionados com o tipo de variável em questão.

Ao se deparar com uma publicação científica, o leitor deve verificar como foram examinados os dados do estudo e a forma de construção de gráficos e tabelas. Falta de clareza na apresentação dos dados e na construção de tabelas é, muitas vezes, indício de que o trabalho deve ser lido com ceticismo e cuidado.

Quando da condução de estudos, o pesquisador deve dar preferência para a coleta das variáveis em suas formas contínuas, sempre que possível. Isto porque, tendo procedido desta forma, fica fácil transformar as variáveis contínuas em categóricas, caso isto seja necessário no momento da análise estatística. O caminho inverso já não é possível, uma vez que variáveis categóricas não podem ser convertidas em contínuas, quando não se dispõe dos valores mensurados em escala contínua. Além desta vantagem na coleta de variáveis contínuas, a análise estatística de dados contínuos apresenta mais poder^b e tende a ser mais simples do que a de dados categóricos, devendo, portanto, ser priorizada.

REFERÊNCIAS

1. Peres KG. Apresentação de dados epidemiológicos. In: Antunes JLF, Peres MA, editores. Fundamentos de odontologia: epidemiologia da saúde bucal. Rio de Janeiro: Guanabara Koogan; 2006. p.409-21.
2. Pereira MG. Epidemiologia: teoria e prática. Rio de Janeiro: Guanabara Koogan; 1995.
3. Altman DG. Practical statistics for medical research. London: Chapman & Hall; 1997.
4. Barbetta PA. Estatística aplicada às ciências sociais. Florianópolis: Editora da UFSC; 2002.
5. Kirkwood BR, Sterne JAC. Essential medical statistics. Oxford: Blackwell Science; 2003.
6. Bastos JL, Gigante DP, Peres KG. Toothache prevalence and associated factors: a population-based study in Southern Brazil. Oral Dis 2007;13:in press.
7. Bastos JL, Peres MA, Peres KG, Barros AJD. Infant growth, development and tooth emergence patterns: a longitudinal study from birth to 6 years of age. Arch Oral Biol 2007;52:in press.

Notas:

^a Os estudos de coorte correspondem a um tipo específico de delineamento na pesquisa epidemiológica. As principais características dos tipos de estudos epidemiológicos serão discutidas em nota a ser publicada futuramente.

^b O poder estatístico corresponde à capacidade de um estudo em identificar uma diferença como estatisticamente significativa quando, de fato, ela o é. Maiores detalhes acerca deste e de conceitos relacionados serão abordados em nota futura.

Endereço para correspondência:
JOÃO LUIZ DORNELLES BASTOS
Avenida do Antão, 353 - Morro da Cruz
CEP 88025-150, Florianópolis, SC, Brasil
Telefone: (48) 3028-1345
E-mail: joao@pilotis.com.br