Centro Universitário São Miguel



Bioestatística

Distribuição de Frequência

[Parte **2** de **2**]

Prof. Me. Yuri Albuquerque





Elementos da distribuição de frequência

Depois de construída uma distribuição de frequência padrão, existem várias outras informações, além da **frequência simples absoluta**, que podem ser incluídas na tabela, as quais auxiliarão na interpretação dos dados. Dentre essas informações, temos:

- ponto médio da classe;
- frequência simples relativa;
- frequência acumulada absoluta; e
- frequência acumulada relativa.





Ponto médio (X_i) ou Valor médio da classe

Corresponde ao ponto equidistante dos limites da classe. Para obter o ponto médio de uma classe, basta somar o limite inferior ao limite superior, e dividir o resultado por dois. Após encontrar o ponto médio da primeira classe, acrescente a amplitude do intervalo de classe para encontrar o ponto médio das classes subsequentes.





Ponto médio (X_i)

Por exemplo: o ponto médio da primeira classe da Tabela é 165 e a amplitude do intervalo de classe é igual a 10, então os pontos médios das classes subsequentes são:

$$X_i = \frac{li + Li}{2} = \frac{160 + 170}{2} = 165$$
, onde: $li =$ Limite inferior; e $Li =$ Limite superior da classe.

Então:
$$165 + 10 = 175$$

$$175 + 10 = 185$$

$$185 + 10 = 195 ...e assim, sucessivamente, para todas as classes.$$

Classes (nº de mL)	Frequência (nº de alunos)
166 - 174	3
175 - 183	4
184 - 192	8
193 - 201	21
202 - 210	14
211 - 219	7
220 - 228	1
229 - 235	2
	$\Sigma f = 60$





Frequência relativa simples (fr)

Refere-se à razão entre a frequência absoluta (f) e o tamanho da amostra (n). Corresponde ao valor que expressa a proporção (ou probabilidade) da ocorrência dos dados de uma classe, de um valor individual ou de uma categoria, em uma tabela de distribuição de frequência. Note que a soma das frequências simples relativas é sempre igual a 1 (Σ fr = 1).

$$fr_i = \frac{f_i}{\sum_{i=1}^k f_i} = \frac{f_i}{n}$$

O resultado também pode ser mostrado em porcentagem (fr%).

 $fr_i\% = \frac{f_i}{n} \times 100$





Frequência acumulada absoluta (F)

De uma classe ou de um valor individual, corresponde à soma da frequência simples absoluta, dessa classe ou desse valor, com as frequências simples absolutas das classes ou dos valores anteriores. É importante saber que a frequência acumulada da última classe é igual ao tamanho da amostra (n).

Frequência acumulada relativa (Fr)

De uma classe ou de um valor individual, corresponde à frequência acumulada dividida por Σf ou n. Também pode ser calculada acumulando as frequências relativas simples de acordo com a definição de frequência acumulada. O resultado também pode ser mostrado em porcentagem (Fr%).



Valores do consumo diário de suco de frutas, por 60 estudantes de uma escola pública

166	170	173	181	182	183	183	184	185	186
186	189	190	190	192	193	193	194	194	194
194	195	195	195	195	195	196	196	197	197
198	198	198	199	201	201	202	202	203	203
204	204	204	205	206	207	207	208	209	210
211	214	215	215	217	218	219	227	229	235





Valores do consumo diário de suco de frutas, por 60 estudantes de uma escola pública

Consumo (mL/dia)	f	Ponto médio	fr	fr (%)	F	Fr	Fr (%)
160 - 170							
170 - 180							
180 - 190							
190 - 200							
200 - 210							
210 - 220							
220 - 230							
230 - 240							
	∑f = 60	_	$\sum \frac{f}{n} = 1$	_			_





Valores do consumo diário de suco de frutas, por 60 estudantes de uma escola pública

Consumo (mL/dia)	f	Ponto médio	fr	fr (%)	F	Fr	Fr (%)
160 - 170	1	165	0,017	1,70	1	0,017	1,70
170 - 180	2	175	0,033	3,30	3	0,050	5,00
180 - 190	9	185	0,150	15,00	12	0,200	20,00
190 - 200	22	195	0,367	36,70	34	0,567	56,70
200 - 210	15	205	0,250	25,00	49	0,817	81,70
210 - 220	8	215	0,133	13,30	57	0,950	95,00
220 - 230	2	225	0,033	3,30	59	0,983	98,30
230 - 240	1	235	0,017	1,70	6	1,00	100,0
	∑f = 60	_	$\sum \frac{f}{n} = 1$	_	_	_	_





Tipos de tabela de distribuição de frequência

Também chamada de distribuição de frequência simples, este tipo de tabela deve ser utilizado na apresentação de dados coletados a partir de uma variável quantitativa, contínua ou discreta, que apresente uma variação relativamente pequena e com um número reduzido de categorias.

Tabela — Distribuição de frequência da idade de 60 alunos matriculados em uma escola pública

Ida(tenos)	f	fr	F	Fr
10	4	0,067	4	0,067
11	9	0,150	13	0,217
12	18	0,300	31	0,517
13	16	0,267	47	0,784
14	8	0,133	55	0,917
15	5	0,083	60	1,000
Σ	60	1,00	_	_





Tabela de grupamento para variáveis categóricas

Utilizadas para variáveis qualitativas, do tipo categórica, essas tabelas são semelhantes às tabelas de frequência simples para variáveis quantitativas, exceto que, na primeira coluna (coluna indicadora), os valores são substituídos por variáveis nominais, as quais devem ser descritas em ordem de frequência decrescente ou crescente, conforme a preferência para exposição dos dados.

Tabela – Distribuição de frequência das espécies de peixe coletadas em um lago

Espécie de peixe	f	%
Tucunaré	436	36,33
Bagre	354	29,50
Pintado	228	19,00
Mapará	124	10,33
Pirarucu	58	4.84
Total	1.200	100,00





Normas para apresentação tabular de dados

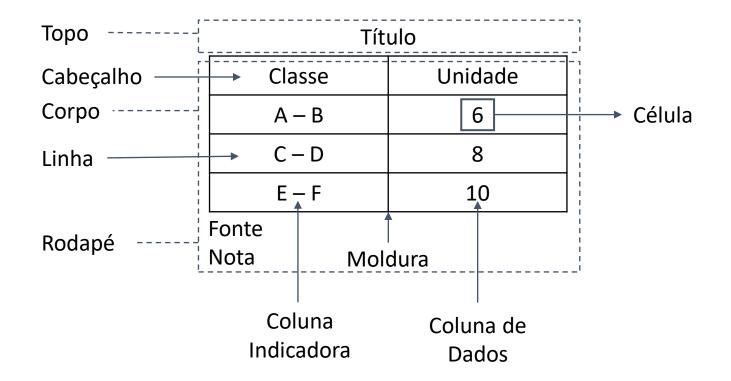
Para que uma tabela seja considerada bem elaborada, é condição fundamental que a mesma seja autoexplicativa. Isto é, que tenha significado próprio, de modo que, quando vista isoladamente, o leitor não precise consultar o texto para entender seu significado e suas informações. Para tanto, é necessário que algumas regras práticas sejam seguidas, as quais são adotadas por órgãos oficiais de estatística.

No Brasil, adotam-se as regras previstas pelo Conselho Nacional de Estatística (CNE), pela Associação Brasileira de Normas Técnicas (ABNT), pelas Normas de Apresentação Tabular do Instituto Brasileiro de Geografia e Estatística (IBGE) e pela Resolução do Conselho Nacional de Metrologia (CONMETRO). A seguir, mostraremos as principais orientações para a elaboração de uma tabela. Para melhores esclarecimentos, o leitor deve consultar as normas acima citadas.





Normas para apresentação tabular de dados





Representação Gráfica

Quadro – Tipos de gráfico quanto à forma

Tipos	Características
• Diagrama	 São gráficos de forma geométrica, dispostos em duas dimensões. Correspondem aos tipos mais utilizados na representação de séries estatísticas.
Cartograma	- São as ilustrações relativas a cartas geográficas, utilizadas para representar dados geográficos, históricos e demográficos.
• Estereograma	- São os gráficos apresentados em três dimensões para a representação de volumes.





Gráficos para dados quantitativos – Histograma de frequências

É o gráfico de colunas mais utilizado para representar as variáveis contínuas. Consiste em um conjunto de retângulos justapostos, cujas bases, localizadas sobre o eixo horizontal, representam as amplitudes dos intervalos de classe, numa escala contínua, e as alturas representam, proporcionalmente, as frequências, absolutas ou relativas das classes. Ao final, o histograma de frequência mostrará uma figura geométrica com área total proporcional à soma de todas as frequências.





Gráficos para dados quantitativos – Histograma de frequências

Pontos importantes:

- Para melhor comparação entre duas distribuições, deve-se utilizar o histograma de frequências relativas ou percentuais.
- Ao empregar as frequências relativas, obtém-se um histograma com área unitária.
- Nas distribuições com classes de intervalos diferentes, é necessário que se faça o ajuste das frequências, para que o gráfico apresente uma figura geométrica proporcional à frequência da ocorrência da variável. Para calcular a altura de cada retângulo, divide-se a sua frequência relativa pela amplitude do intervalo de sua classe (h). Se todas as classes apresentarem amplitudes iguais, não é necessário realizar a divisão.

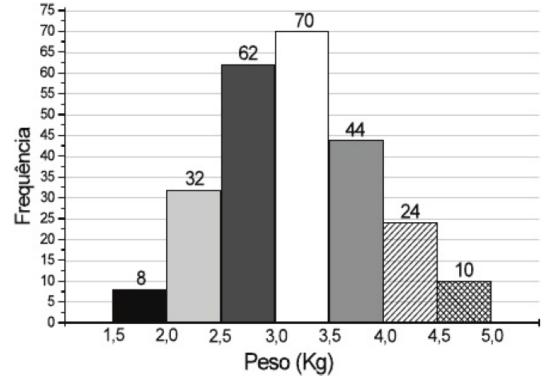




Histograma de frequências

Peso ao nascer de 250 crianças nascidas vivas na maternidade A, em 2008

Peso(kg)	Frequência
1,5 2,0	8
2,0 2,5	32
2,5 3,0	62
3,0 3,5	70
3,5 4,0	44
4,0 4,5	24
4,5 l— 5,0	10
Total	250



Histograma de frequência do peso ao nascer de 250 crianças nascidas vivas na maternidade A, em 2008





Polígono de frequência

Menos utilizado que o gráfico de colunas, o polígono de frequência é um gráfico de linha originado a partir de uma distribuição de frequência. Essa linha poligonal fechada, quase sempre traçada acompanhando um histograma, une as ordenadas marcadas pelos pontos médios das classes do respectivo histograma. Para construir um polígono de frequência, utilizam-se as mesmas escalas horizontal e vertical que foram empregadas para o histograma, só que marcadas com o ponto médio das classes. Então, marcam-se os pontos que correspondem ao ponto médio e à frequência de cada classe, os quais são conectados por uma linha traçada da esquerda para a direita. Uma vez que o gráfico deve mostrar um polígono fechado, a linha deve começar no eixo horizontal, a qual deve ser estendida, para a esquerda, em uma amplitude de classe, antes do ponto médio da primeira classe, e, estendida, de igual modo, para a direita, depois da última classe.

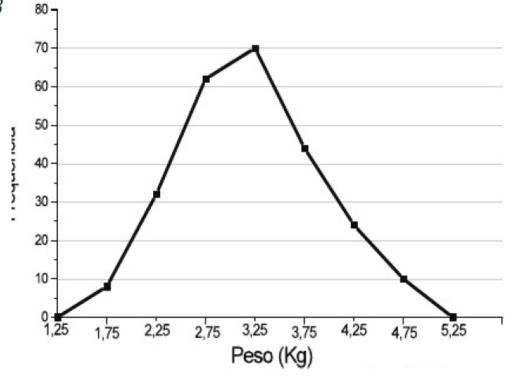




Polígono de frequência

Peso ao nascer de 250 crianças nascidas vivas na maternidade A, em 2008

Peso(kg)	Frequência
1,5 l— 2,0	8
2,0 2,5	32
2,5 l— 3,0	62
3,0 3,5	70
3,5 4,0	44
4,0 4,5	24
4,5 I— 5,0	10
Total	250



Polígono de frequência do peso ao nascer de 250 crianças nascidas vivas
 na maternidade A, em 2008





Polígono de frequência acumulada (ogiva) ou ogiva de Galton

É um gráfico de linha que mostra a frequência acumulada de cada classe, em seu limite superior. Os limites superiores são marcados no eixo horizontal e as respectivas frequências acumuladas são marcadas no eixo vertical. Após marcar os pontos, traça-se a linha conectando-os da esquerda para a direita. O gráfico poderá iniciar no limite inferior da primeira classe, onde a frequência acumulada é zero, e terminar no limite superior da última classe, onde a frequência acumulada é igual ao tamanho da amostra.

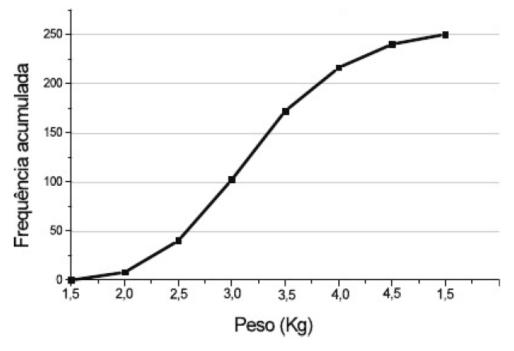




Polígono de frequência acumulada (ogiva) ou ogiva de Galton

Peso ao nascer de 250 crianças nascidas vivas na maternidade A, em 2008

Peso(kg)	Frequência
1,5 2,0	8
2,0 I— 2,5	32
2,5 3,0	62
3,0 3,5	70
3,5 4,0	44
4,0 4,5	24
4,5 l— 5,0	10
Total	250



Polígono de frequência acumulada do peso ao nascer de 250 crianças nascidas vivas na maternidade A, em 2008





Gráfico de caule e folha (steam-and-leaf)

Desenvolvido por John Tukey em 1977, o diagrama de caule e folha é um método atual de apresentação de dados quantitativos, especialmente quando esses dados contêm muitos valores ou diferentes categorias, o que os torna difíceis de serem agrupados em tabelas e representados graficamente. Este tipo de gráfico, que mostra os valores dispostos em forma de uma curva, facilitando a visualização da sua simetria e curtose, apresenta a vantagem de mostrar os dados em sua forma original, sem que nenhuma informação seja perdida, exceto pela ordem com que foram coletados, além de facilitar o seu ordenamento. Para construir o gráfico, cada número da distribuição é separado em um caule, que corresponde a todos os seus dígitos, menos o dígito mais à direita, o qual corresponde à folha.





Gráfico de caule e folha (steam-and-leaf)

147	126	129	155	168	168	142	159	145	116
130	122	112	126	118	122	109	140	120	117
110	00	100	110	139	122	133	145	134	132
110	09	100	100	100	140	130	108	118	98
	148	106	120	138	140	109	100		U

Escores	Caule	Caule Folhas									
1	8	9									
1	9	8									
3	10	8	9	9		1362					
7	11	2	6	7	8	8	8	9			
9	12	2	2	2	6	6	6	6	9	9	
8	13	0	2	3	3	4	8	9	9		
7	14	0	0	2	5	5	7	8		COR	
2	15	5	9					SUR		expe	
2	16	8	8								

g. 7.4 – Gráfico de caule e folhas dos valores da pressão arterial sistólica de 40 pacientes aleatoriamente selecionados





Gráfico de distribuição conjunta ou diagrama de dispersão

Também chamado de diagrama de dispersão, o gráfico de distribuição conjunta é utilizado para representar e agrupar dados que envolvam duas variáveis quantitativas, simultaneamente, de tal maneira que os dois conjuntos de dados formem pares ordenados, com cada um desse pares pertencente ao mesmo sujeito da pesquisa. Portanto, o objetivo do gráfico é mostrar se existe uma associação entre as duas variáveis estudadas, assim como identificar o tipo de correlação entre elas, se positiva ou negativa, de tal modo que se possa predizer o comportamento de uma em função da variação da outra. É o tipo de gráfico utilizado nos testes de correlação de Pearson e regressão linear. O gráfico é construído marcando-se os valores de uma das variáveis, preferencialmente a preditora, sobre o eixo horizontal (eixo x), e os valores da outra variável, a variável-resposta, sobre o eixo vertical (eixo y). Ao final, obtém-se um gráfico cartesiano com múltiplos pontos, cada um deles representando um par ordenado (x e y).





Gráfico de distribuição conjunta ou diagrama de dispersão

Valores das pressões arteriais sistólica e diastólica de uma amostra de 15 indivíduos

Indivíduo	Pressão arterial (mmHg)		
IIIulviuuo	Sistólica	Diastólica	
01	108	62	
02	134	74	
03	100	64	
04	108	68	
05	112	72	
06	112	64	
07	112	68	
08	122	70	
09	116	70	
10	116	70	
11	120	72	
12	108	70	
13	108	70	
14	114	74	
15	108	68	

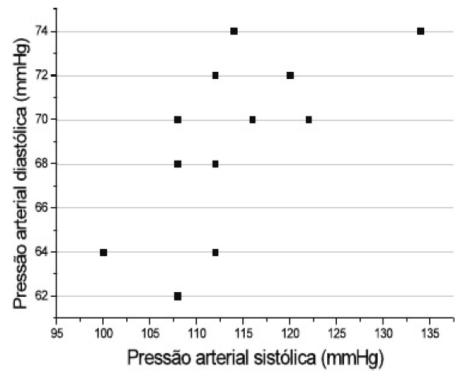


 Gráfico de distribuição conjunta dos valores das pressões arteriais sistólica e diastólica de 15 indivíduos





Gráficos para dados qualitativos – Gráfico de colunas

É um gráfico utilizado para comparar grandezas de diferentes categorias de uma mesma variável. É formado por retângulos separados, verticalmente dispostos, cujas bases, de mesma largura, situam-se no eixo horizontal, e cujas alturas são proporcionais às grandezas que representam. Os dados podem ser apresentados em forma de números absolutos ou em proporções.





Gráficos para dados qualitativos – Gráfico de colunas

Exemplo de distribuição qualitativa

Conceitos	Frequência
Senconceito	2
Insuficiente	8
Regular Bom	16 10
Excelente	4
Total	40

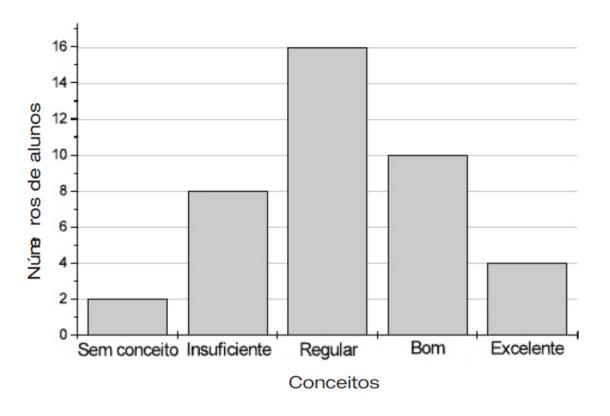


Gráfico de colunas dos conceitos de 40 alunos da turma de bioestatística





Gráfico de colunas com barra de erros

As barras de erros são traços em forma de "T," situados no topo dos retângulos do gráfico de colunas. Devem ser utilizadas quando a altura dos retângulos representa a média aritmética dos valores das categorias da variável mostrada no estudo. A critério do autor, as barras de erro podem significar o desvio padrão ou o erro padrão da média da variável apresentada e têm o propósito de fornecer uma informação visual para que o leitor possa comparar os retângulos e verificar qual ou quais deles são estatisticamente diferentes. Sempre que possível, estas barras devem estar presentes nos gráficos de colunas, de barras ou de linhas, as quais têm os seus comprimentos calculados pelo software empregado para a confecção do gráfico.

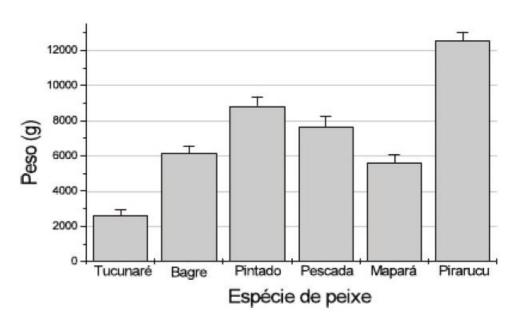




Gráfico de colunas com barra de erros

Média ± desvio padrão dos pesos decinco espécies de peixes daregião amazônica

Espéciedepeixe	Peso±DP (g)
Tucunaré	2.630±316
Bagre	6.140 ± 436
Pintado	8.850±528
Pescada	7.640±620
Mapará	5.620 ± 432
Pirarucu	18.580±858



^{7 –} Gráfico de colunas com barras de erro dos pesos de seis espécies de peixes





Diagrama de pareto

É uma variante do grá©co de colunas que ordena as frequências das ocorrências, da maior para a menor, permitindo melhor visualização e interpretação dos dados apresentados. Mostra ainda a curva de percentagens acumuladas. Tem como base o princípio de Pareto 3, que diz: alguns elementos são vitais; muitos, apenas triviais. Há muitos problemas sem importância diante de outros mais graves. Este princípio também ficou conhecido como "Lei 20/80".





Diagrama de pareto

Valores gastos, em milhões de reais, conforme a especialidade médica, no

hospital A - 2008

Especialidade médica	Gastos (R\$)	Gastos (%)	Acumulada (%)
Cirurgia	58,32	32,40	32,40
Ortopedia	50.94	28,30	60,70
Ginecologia	28,98	16,10	76,80
Pneumologia	21,42	11,90	88,70
Clínicamédica	13,32	7,40	96,10
Pediatria	7,02	3,90	100,00
Total	180		

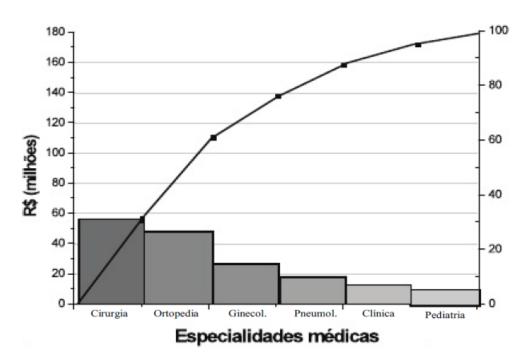


Diagrama de Pareto dos gastos de um hospital público, conforme a especialidade médica





Gráfico de colunas justapostas

Também chamado de gráfico comparativo, o gráfico de colunas justapostas deve ser utilizado quando o objetivo é descrever, simultaneamente, duas ou mais categorias, para uma única variável, em diferentes amostras. Neste gráfico, as colunas que representam as categorias, são colocadas lado a lado, de modo que se possa compará-las quanto à grandeza que elas representam.





Gráfico de colunas justapostas

Resultado do teste de avaliação de conhecimento, aplicado em duas escolas

públicas do segundo grau

Dissiplins	Escola		
Disciplina	Estadual	Municipal	
Matemática	67 4 9	645	
Física	6 2 4	596	
Química	7345	78 <u>4</u> 6	
Biologia	8748	8 <u>14</u> 5	

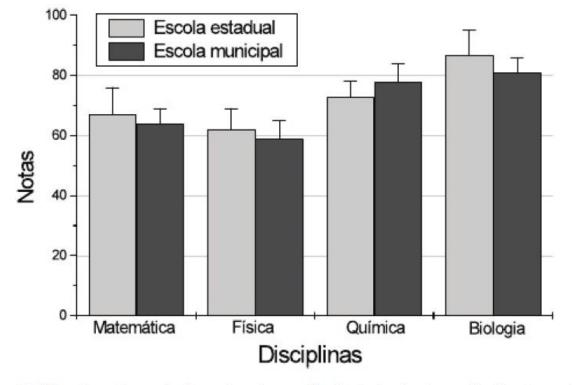


Gráfico de colunas justapostas do resultado do teste de avaliação de conhecimento, aplicado em duas escolas públicas do segundo grau





Gráfico de colunas superpostas

Também chamado de gráfico de colunas compostas, igualmente ao gráfico de colunas justapostas, serve para comparar, simultaneamente, duas ou mais categorias, para uma única variável, em diferentes amostras. Neste gráfico, as colunas que representam as categorias são colocadas uma sobre a outra, de modo que se possa compará-las quanto à grandeza que elas representam.





Gráfico de colunas superpostas

Relação entre as diferentes partes do corpo e o peso corporal total, de seis espécies de peixe da região amazônica

Ecnópio do	Partes do corpo (%)		
Espécie de peixe	Cefálica	Corpo	Nadadeira caudal
Tucunaré	14,0	82,0	4,0
Bagre	29,0	66,0	5,0
Pintado	23,0	73,0	4,0
Pescada	17,0	77,0	6,0
Mapará Pirarucu	25,0 18,0	70,0 78,0	5,0 4,0

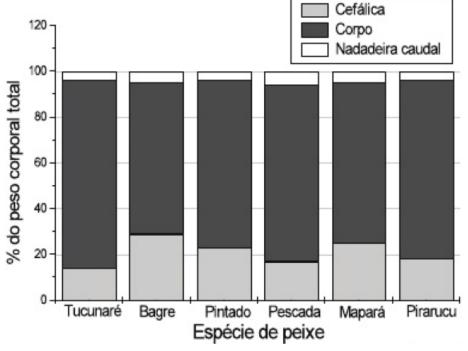


Gráfico de colunas superpostas da relação entre as diferentes partes do corpo e o peso corporal total, de seis espécies de peixe da região amazônica





Gráfico de setores

Também chamado de gráfico de pizza ou torta, em razão da sua aparência circular, este tipo de gráfico deve ser utilizado quando se deseja apresentar dados nominais ou ordinais, cujo principal objetivo é mostrar a relação entre as partes e o todo. A representação é feita tomandose como base a figura de um círculo de raio qualquer, o qual tem o seu ângulo central proporcionalmente dividido em setores. Para construí-lo, devem ser calculadas as frequências relativas, ou porcentagens, de cada categoria. Então, calcula-se o ângulo central de cada setor multiplicando-se 360o pela frequência relativa da categoria.





Gráfico de setores

7 – Tipos de cirurgias realizadas no hospital A, em 2008

Tipodecirurgia	f	fr%
Abdominal	1.008	42,0
Ortopédica	552	23,0
Ginecológica	432	18,0
Torácica Pediátrica	240 96	10,0 4,0
Plástica	72	3,0
Total	2.400	

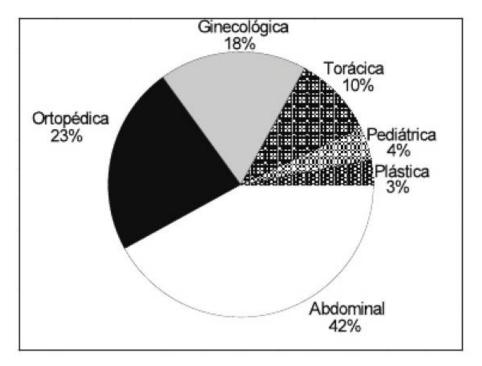


Gráfico de setores dos tipos de cirurgias realizadas no hospital A, em 2008





Gráfico de linha (curva)

Este tipo de gráfico geralmente é utilizado para a apresentação de séries temporais, onde o conjunto de dados é composto por valores tomados em intervalos regulares. O eixo horizontal representa o intervalo de tempo, ao passo que o eixo vertical marca a evolução dos valores para a variável em estudo. Após marcar todos os pontos, os pares ordenados são, então, unidos por segmentos de linha. É possível, em um mesmo gráfico, representar duas ou mais variáveis, simultaneamente.





Gráfico de linha (curva)

Valores relativos aos gastos com projetos de pesquisa, em três universidade s públicas

Ano	Gastos por universidade (em milhões)		
3	Α	В	С
2000	1,2	4,8	7,9
2001	3,5	6,8	9,8
2002	4,7	8,7	15,1
2003	7,6	13,8	19,8
2004	11,4	17,7	25,8
2005	14,8	22,9	32,9
2006	18,5	29,9	58,8
2007	26,8	47,9	67,9
2008	42,1	58,8	74,2

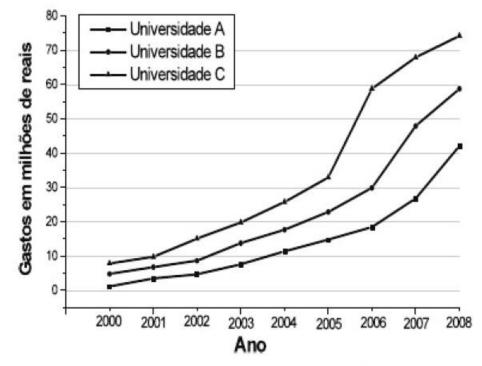


Gráfico de linha dos valores relativos aos gastos com projetos de pesquisa, em três universidades públicas



REFERÊNCIAS

- Fontelles, Mauro José. Bioestatística: aplicada à pesquisa experimental. São Paulo: Livraria da Física, vol 1, 2012.
- Vieira, Sonia. Introdução à Bioestatística. 6º Ed. Rio de Janeiro: GEN, 2021.

ONTEÚDO DA AULA ONTEÚDO DA AUL

CONTATOS









