

População, amostra, variável, coleta de dados, apuração de dados e apresentação tabular.

Estatística: fornece uma coleção de métodos para planejar experimentos, para obter e organizar dados, resumi-los, analisá-los, interpretá-los e deles extrair conclusões (Triola, 1999).

Bioestatística – Estatística aplicada às ciências da vida.

Níveis de mensuração

Escala nominal

Os indivíduos (ou unidades de análise) são classificados em categorias segundo uma característica.

Exemplos

Sexo (masculino, feminino),

Hábito de fumar (fumante, não fumante),

Sobrepeso (sim, não),

Condição do domicílio (próprio já pago, próprio em pagamento, alugado, cedido, outra condição);

Sexo dos insetos vetores (fêmea, macho);

Tipo de habitat (mata, margem da mata, campo aberto, domicílio);

Local do domicílio (intradomicilio e peridomicilio)

Característica:

Não existe ordem entre as categorias e suas representações, se numéricas, são destituídas de significado numérico.

Ex: Sexo do paciente
Masculino =1, feminino = 2
Os valores 1 e 2 são apenas rótulos

Ex: Sexo do inseto
Fêmea=1, Macho = 2
Os valores 1 e 2 são apenas rótulos e não podem ser tratados como números.

Tipo de habitat de determinado inseto
1= mata, 2 = margem da mata, 3= campo aberto, 4= domicílio
Da mesma forma, os valores 1, 2, 3, e 4 são apenas rótulos.

Escala ordinal

Os indivíduos são classificados em categorias que possuem algum tipo inerente de ordem. Neste caso, uma categoria pode ser "maior" ou "menor" do que outra.

Ex: Nível socioeconômico (A, B, C e D; onde A representa maior poder aquisitivo);

Nível de retinol sérico (alto, aceitável, baixo, deficiente) critérios: *Committee on Nutrition for National Defense ICNND/USA*, 1963 (in Prado MS et al, 1995).

Alto: maior ou igual a 50,0 µg/dl;

Aceitável: 20,0 a 49,9 µg/dl;

Baixo: 10,0 a 19,9 µg/dl;

Deficiente: menor ou igual a 10,0 µg/dl.

Tamanho da asa de mosquitos culicídeos classificados em categorias (Landry, SV et al., 1988, *Journal of the American Mosquito Control Association*- vol.4 nº 2).

Pequeno ($\leq 2,00$ mm),

Médio (2,01 – 3,64 mm),

Grande ($\geq 3,65$ mm).

Embora exista ordem entre as categorias, a diferença entre as categorias adjacentes não tem o mesmo significado em toda a escala. Se as categorias forem representadas por números, estas são destituídas de significado numérico.

Escala numérica intervalar

Este nível de mensuração possui um valor zero arbitrário e, por este motivo não permite calcular a razão entre dois valores, sendo possível, entretanto calcular a soma e subtração.

Como exemplo deste nível de aferição temos a temperatura em graus Celsius e *Fahrenheit*. O exemplo abaixo indica o efeito do zero arbitrário na utilização de operações matemáticas (diferença e divisão) tanto em uma variável aferida pela escala numérica intervalar como por uma em escala de razões contínua.

A variável de aferição é temperatura em graus Celsius.

material	$^{\circ}\text{C}$	$^{\circ}\text{F}$	dif $^{\circ}\text{C}$	dif $^{\circ}\text{F}$	dif $^{\circ}\text{C}$ /dif $^{\circ}\text{F}$	razão $^{\circ}\text{C}$	razão $^{\circ}\text{F}$	Razão $^{\circ}\text{C}$ /razão $^{\circ}\text{F}$
A	20	68	A-B =20	A-B =36	0,56	A/B=0,50	A/B=0,65	0,77
B	40	104	B-C =20	B-C =36	0,56	B/C=0,67	B/C=0,74	0,91
C	60	140	A-C =40	A-C =72	0,56	A/C=0,33	A/C=0,49	0,67

A variável de aferição é comprimento (cm) em escala de razões contínua

comprimento	cm	polegada	difcm	dif pol	Difcm/difpol	Razão cm	Razão pol	Razão cm /razão pol
A	20	50,8	A-B =15	A-B =38,1	0,394	A/B=0,571	A/B=0,571	1
B	35	88,9	B-C =5	B-C =12,7	0,394	B/C=0,875	B/C=0,875	1
C	40	101,6	A-C =20	A-C =50,8	0,394	A/C=0,5	A/C=0,5	1

Escala (numérica) de razões discretas

A escala de razões possui zero inerente de acordo com a natureza da característica sendo aferida. No caso de razões discreta, o resultado numérico da aferição é um valor inteiro. Normalmente trata-se de contagem.

Exemplos

Número de refeições em um dia (nenhuma, uma, duas, três, quatro, ...),

Frequência de consumo semanal de determinado alimento (1 vez, 2 vezes, 3 vezes, 4 vezes, 5 vezes, 6 vezes, 7 vezes),

Número de exemplares de *Aedes aegypti* na forma imatura (larvas, pupas)

Número de exemplares capturados (2, 3, 10, 30, 40, 50, 100 ...)

Número de ovos postos (1, 2, 20, 30, ... 50, 100).

Quantidade de repastos sanguíneos realizados por fêmea de inseto (1, 2, 3, 4, 5)

Escala de razões contínua

O resultado numérico é um valor pertencente ao conjunto dos números reais $R = \{-\infty; \dots; 0; 0,2; 0,73; 1; 2,48; \dots; +\infty\}$.

Exemplos

Idade (anos)
 Peso (g)
 Altura (cm)
 Nível de retinol sérico ($\mu\text{g/dl}$)
 Circunferência da cintura (cm)
 Precipitação pluviométrica em mm^3 (quantidade de chuva por metro quadrado)
 Tamanho da asa de um inseto (mm)
 Peso seco de fêmeas ou pupas de mosquitos (0,53 mg, 0,43 mg;...)
 Volume do repasto sanguíneo em (μl) (4,7; 3,6; 4,0; 4,9 ...)
 Comprimento da asa em (mm)
 Largura de partes do corpo de Triatomíneos em (mm)

Tipos de variáveis

De acordo com os níveis de mensuração, pode-se classificar a **natureza das variáveis** segundo a escala de mensuração em:

VARIÁVEL:	{	qualitativa	{	nominal
				ordinal
		quantitativa	{	discreta
				contínua

O tipo da variável irá indicar a melhor forma para o dado ser apresentado em tabelas e gráficos, em medidas de resumo e, a análise estatística mais adequada.

TABLE 1
Neurological manifestations of dengue in a case series of patients recruited in Central Brazil, 2005-2006.

Patient/ number	Sex	Age (years)	Fever/ days	2009 WHO classification	Neurological/ manifestation	MAC-ELISA	Virus isolation	Molecular test	Outcome
1	F	52	4	Dengue	Paresthesia/hands	Pos	Neg	DENV-3	Cured
2	F	3	5	Dengue	Paresthesia/face	Pos	Neg	---	Cured
3	F	35	6	DWS	Paresthesia/UL+LL	Pos	Neg	---	Cured
4	M	36	3	Severe	Paresthesia/hands	Pos	DENV-3	DENV-3	Cured
5	F	17	5	DWS	Paresthesia/lips	Pos	Neg	---	Cured
6	F	56	8	Severe	Encephalo	Pos	Neg	---	Cured
7	F	24	5	Severe	Paresthesia/UL+LL	Undeter.	DENV-3	DENV-3	Cured
8	F	20	6	DWS	Paresthesia/LL	Pos	Neg	---	Cured
9	F	31	9	DWS	Paresthesia/LL	Pos	Neg	---	Cured
10	F	36	7	DWS	Paresthesia/hands	Pos	Neg	---	Cured
11	F	29	7	Dengue	Paresthesia/hands	Pos	Neg	---	Cured
12	F	47	6	Dengue	Paresthesia/hands	Pos	Neg	---	Cured
13	F	41	10	Severe	Encephalo /Seizures	Pos	Neg	DENV-3	Death
14	F	45	6	Dengue	Paresthesia/LL	Pos	Neg	---	Cured
15	F	25	7	Severe	Paresthesia/LL	Pos	Neg	---	Cured
16	F	17	8	Severe	Meningoenceph	Pos	Neg	---	Cured
17	F	19	1	DWS	Paresthesia/feet	Pos	Neg	---	Cured
18	F	54	7	Dengue	Paresthesia+Paresis	Pos	Neg	---	Cured
19	F	27	6	DWS	Paresthesia	Pos	Neg	---	Cured
20	F	34	7	DWS	Paresthesia/LL	Pos	Neg	---	Cured
21	M	68	10	Severe	Encephalo	Pos	Neg	---	Cured
22	F	39	3	DWS	Paresthesia/UL+LL	Pos	Neg	---	Cured
23	M	14	5	Severe	Encephalo	Pos	Neg	---	Cured
24	M	71	15	Severe	Encephalitis/Paresis	Pos	Neg	---	Cured
25	F	38	7	DWS	Paresthesia/hands	Pos	Neg	---	Cured
26	F	15	16	Severe	Meningoenceph	Pos	Neg	DENV-3	Death
27	M	24	1	Severe	Encephalo/Seizures	Pos	Neg	---	Cured
28	F	18	3	Severe	Seizures	Pos	Neg	---	Cured

WHO: World Health Organization; MAC-ELISA: immunoglobulin M (IgM) antibody-capture enzyme-linked immunosorbent assay; F: female; M: male; DENV: dengue virus; DWS: dengue warning sign; LL: lower limbs; UL: upper limbs; Severe: severe dengue; Encephalo: encephalopathy; Meningoenceph: meningoencephalitis; Pos: positive; Neg: negative

Banco de dados para estudar e treinar

Considerar a pesquisa realizada em 2013, com 50 idosos do município de São Paulo. Entre as características investigadas foram obtidos dados do sexo do participante, peso e altura para construção do índice de massa corporal (imc); perguntou-se sobre doenças crônicas não transmissíveis (diabetes, hipertensão, doenças respiratórias e outras doenças crônicas) registrando-se o número de doenças no momento da pesquisa e nível de triglicérides (mg/dL).

id	idade	sexo	doenças crônicas	imc	triglic	id	idade	sexo	doenças crônicas	imc	triglic
1	94	M	1	26	128	26	82	F	1	24	89
2	74	F	4	31	166	27	82	F	1	34	92
3	74	F	1	24	79	28	85	F	4	25	181
4	64	F	0	22	166	29	87	F	3	20	91
5	61	F	2	27	61	30	74	F	3	27	171
6	89	F	0	27		31	72	F	3	45	176
7	84	F	3	26	211	32	83	F	3	35	165
8	73	M	2	27	157	33	91	F	1	24	38
9	93	F	1	28	124	34	73	F	1	22	46
10	87	F	3	26	111	35	66	F	1	31	
11	83	M	0	24	80	36	82	F	2	27	153
12	78	M	2	27	73	37	82	M	3	23	
13	76	M	1	23	205	38	85	F	2	20	99
14	76	F	1	29	101	39	86	F	2	29	66
15	72	M	3	24		40	92	M	3	29	130
16	65	F	2	35	170	41	71	M	6	27	72
17	68	M	2	29	126	42	75	M	0	30	87
18	66	F	1	37	193	43	74	M	1	34	219
19	91	M	0	19	92	44	61	M	0	25	
20	89	M	1	23	47	45	64	F	2	34	125
21	78	F	3	19	221	46	62	F	4	29	233
22	93	F		28	86	47	80	F	2	27	118
23	71	M	0	28	119	48	80	F	3	23	56
24	88	F	3	26	75	49	91	F	2	29	80
25	80	F	2	28	145	50	86	F	3	27	104

Exercício

Classificar quanto à natureza, as variáveis

Idade:

Sexo:

Doenças crônicas:

IMC:

Triglicérides:

A característica (variável) imc pode ser utilizada em categorias, por exemplo
abaixo de 21 indicando magreza;
de 22 a 27 eutrofia e
28 e mais, excesso de peso

Exercício

Classificar quanto à natureza, as seguintes variáveis:

Variável	Tipo (natureza)
Condição de saúde (doente, não doente)	
Tipo de parto (normal, cesáreo)	
Nível de colesterol sérico (mg/100cc)	
Tipo de abrigo (intradomiciliar, peridomiciliar)	
Estado nutricional (desnutrição, eutrofia, sobrepeso, obesidade)	
Altitude da área de coleta (metros)	
Resultado sorológico para presença de vírus (reagente, não reagente)	
Número de larvas e pupas em determinado criadouro	
Peso seco de pupas de exemplares de <i>Aedes aegypti</i> (mg)	
Estado de paridade da fêmea de inseto vetor (nulípara e parida)	
Tempo de um procedimento cirúrgico (minutos)	
Número de praias consideradas poluídas	

Coleta de dados

É a observação e registro das categorias ou das medidas das variáveis relacionadas ao objeto de estudo que ocorrem em unidades (indivíduos) de uma amostra ou população.

Definições e notação

População: totalidade de elementos sob estudo. Apresentam uma ou mais características em comum.

Supor o estudo sobre a ocorrência de mosquitos vetores de malária no Parque Estadual da Serra da Cantareira, município de São Paulo.

População alvo – larvas de anofelinos do subgênero *Kerteszia*

População de estudo – larvas do gênero *Anopheles* subgênero *Kerteszia* que se criam em bromélias na trilha do Pinheirinho fixadas em até 15 metros de altura e que estejam em condições de identificação.

Supor o estudo sobre a ocorrência de sobrepeso em crianças de 7 a 12 anos no Município de São Paulo.

População alvo – todas as crianças nesta faixa etária deste município.

População de estudo – crianças matriculadas em escolas.

Elementos: são unidades de análise por exemplo pessoas, células, domicílios, armadilhas, bromélias ou outro tipo de criadouro.

Amostra: é uma parte da população de estudo.

Amostragem: processo para obtenção de uma amostra. Tem como objetivo estimar parâmetros populacionais.

Parâmetro: Quantidade fixa de uma população.

Ex: Quantidade média de sangue ingerido por uma fêmea de mosquito, em uma picada.

Temperatura média no processo de transformação de larva em pupa.

Estimador: é uma fórmula matemática que permite estimar um parâmetro. Pode ser estimador no ponto e por intervalo.

Ex: Estimador no ponto

$$\text{Média aritmética: } \bar{X} = \frac{\sum_{i=1}^N X_i}{N},$$

$$\text{onde } \sum_{i=1}^N X_i = X_1 + X_2 + \dots + X_N \text{ e } N = \text{número de observações.}$$

Estimador por intervalo



Tabela 1 - Prevalência (%) de sedentarismo no lazer e global segundo variáveis socioeconômicas e demográficas em homens adultos em áreas do Estado de São Paulo, Brasil.



	HOMEM				
	N	Inativos no lazer		Inativos IPAQ	
		Prevalência (%)	Razão de prevalência (IC 95%)	Prevalência (%)	Razão de prevalência (IC 95%)
Faixa etária					
18 a 29	474	44,7	1	19,8	1
30 a 39	204	59,0	1,10 (1,03-1,17)	19,4	1,00 (0,92-1,08)
40 a 49	198	64,9	1,14 (1,07-1,21)	29,0	1,08 (0,98-1,19)
50 a 59	144	65,2	1,14 (1,06-1,23)	28,5	1,07 (0,99-1,16)
Total	1020	56,2		23,4	
		p=0,000		p=0,150	
Cor*					
Branca	716	54,1	1	26,4	1
Preta/parda	281	61,8	1,05 (1,00-1,10)	16,7	0,92 (0,87-0,98)
		p=0,050		p=0,016	

Estimativa: Valor do estimador calculado em uma amostra. Estima o valor do parâmetro.

Ex: Peso seco médio (mg) de fêmeas de *Culex quinquefasciatus*: $\bar{x} = 0,541$ mg. Os pesos (mg) para cada indivíduo eram: 0,419; 0,641; 0,592; 0,477; 0,613; 0,501.

Ex: Peso médio ao nascer, calculado em uma amostra de 120000 crianças nascidas no Município de São Paulo no ano de 2000: média amostral = $\bar{x} = 3000$ g .

Indicações para utilizar uma amostra

População muito grande
 Processo destrutivo de investigação
 Novas terapias

Vantagens de realizar um estudo com amostragem:

Menor custo
 Menor tempo para obtenção dos resultados
 Possibilidade de objetivos mais amplos
 Dados possivelmente mais fidedignos

Desvantagens

Resultados sujeitos à variabilidade

Se fossem retiradas amostras de uma população e calculada, por exemplo, a média, os valores das médias (estimativas) não seriam necessariamente iguais. Abaixo segue um exemplo de amostras de tamanho 5 retiradas dos dados do exercício 1, para a variável idade.

Amostra 1	Amostra 2	Amostra 3
76	74	94
82	86	76
65	72	65
84	73	72
71	87	83
$\bar{x}=75,6$ anos	$\bar{x}=78,4$ anos	$\bar{x}=78,0$ anos

Considerando-se os 50 idosos como a população, a média de idade (μ) é **78,7 anos** (soma-se todos os valores e divide-se o resultado por 50)

Tipos de Amostragem

Probabilística: cada unidade amostral tem probabilidade conhecida e diferente de zero de pertencer à amostra. É usada alguma forma de sorteio para a obtenção da amostra.

Não probabilística: não se conhece a probabilidade de cada unidade amostral pertencer à amostra. Algumas unidades terão probabilidade zero de pertencer à amostra.

Ex: amostragem intencional; por voluntários; acesso mais fácil; por quotas.

Tipos de amostragem probabilística:

- aleatória simples (com e sem reposição);
- sistemática;
- com partilha proporcional ao tamanho do estrato;
- por conglomerado.

Amostragem aleatória simples (AAS)

É o processo de amostragem onde qualquer subconjunto de n elementos diferentes de uma população de N elementos tem mesma probabilidade de ser sorteado. Tamanho da população: N ; tamanho da amostra: n ; fração global de amostragem ou probabilidade de sortear um

$$\text{indivíduo} = \frac{n}{N}.$$

- É necessário ter um sistema de referência que contenha todos os elementos da população da qual será retirada a amostra;
- Utilização da tabela de números aleatórios – mecânica;
- Utilização de programas computacionais.

Considerações a respeito da tabela de números aleatórios

Para a utilização desta tabela é necessário:

- Definir o número de dígitos que serão utilizados
- Sortear um início
- Pré definir um caminho a ser percorrido

Tabela de números equiprováveis (aleatórios)

61	09	26	29	85	11	95	77	79	04	57	00	91	29	59	83	53	87	02	02
94	47	40	99	93	82	13	22	40	33	19	72	55	69	82	16	94	21	66	39
50	40	50	55	79	00	58	17	26	30	38	11	54	89	04	13	69	17	35	48
51	01	75	76	54	43	11	28	32	75	33	09	04	78	74	91	56	79	43	39
25	45	79	30	63	56	44	70	05	04	31	81	46	02	92	32	06	71	12	48
63	94	61	14	24	60	27	00	00	95	54	31	59	00	79	94	46	32	61	90
12	95	04	73	06	72	76	88	55	62	38	79	18	68	10	31	93	58	66	92
38	06	78	00	85	42	57	29	28	34	79	91	93	58	82	97	37	07	64	67
22	69	28	18	25	08	90	93	53	17	54	12	21	03	56	30	88	53	46	82
07	95	63	14	76	53	62	10	21	57	55	74	57	68	22	38	84	55	57	49
61	41	81	16	97	55	19	65	08	62	26	38	74	32	30	44	64	64	91	80
97	15	71	92	40	28	33	35	23	32	75	36	18	98	41	10	50	93	75	95
39	81	34	84	33	83	42	77	35	00	51	42	82	63	30	47	01	98	96	73
58	35	04	52	06	81	24	32	74	53	28	82	43	35	01	73	34	47	05	76
52	85	30	59	37	00	49	88	07	43	08	04	00	48	36	23	31	88	80	88
41	92	93	01	94	13	33	63	32	35	38	91	18	89	71	67	46	73	42	47
88	51	22	59	99	51	20	74	13	55	30	41	25	99	10	26	01	33	24	13
11	12	32	28	25	67	22	97	11	73	55	24	09	23	47	12	93	44	80	47
33	02	06	80	29	39	78	49	81	21	42	00	99	80	44	56	33	83	46	16
03	67	08	29	16	04	92	31	62	03	94	53	02	60	55	72	46	68	25	93
41	54	93	90	86	52	14	58	90	34	83	00	73	38	14	50	77	58	08	94
18	84	83	61	42	96	82	86	02	30	40	16	65	55	63	20	40	24	79	80
06	15	93	11	72	17	32	31	84	89	53	66	01	99	53	75	79	92	20	61
12	74	92	15	60	93	84	37	29	62	24	96	78	93	28	34	41	69	04	51
79	13	36	81	55	51	46	66	68	85	07	73	35	42	52	61	29	21	02	34
01	78	33	32	06	16	45	94	09	18	40	14	73	03	61	80	69	79	52	95
90	73	28	21	38	57	39	36	24	33	31	99	64	86	19	61	55	50	65	14
44	10	20	96	70	32	41	46	22	97	08	22	02	47	43	57	15	87	76	59
52	47	00	27	41	43	70	17	52	44	51	26	94	73	17	72	16	51	81	77
23	03	84	44	29	43	57	05	46	59	89	00	65	01	20	27	32	66	34	56

Amostragem sistemática

Utiliza-se a ordenação natural dos elementos da população (prontuários, casa, ordem de nascimento).

- Intervalo de amostragem $k = \frac{N}{n}$, onde
N= tamanho da população e n = tamanho da amostra
- Início casual i, sorteado entre 1 e k, inclusive
- Amostra sorteada é composta pelos elementos: i, i+k, i+2k, ..., i+(n-1)k

OBS: É necessário ter cuidado com a periodicidade dos dados, por exemplo se for feito sorteio de dia no mês, pode cair sempre em um domingo onde o padrão de ocorrência do evento pode ser diferente.

Exemplo: N=80; n=10; $k = \frac{N}{n} = \frac{80}{10} = 8$; início casual: $1 \leq i \leq 8$

Começo casual **sorteado**: i=4

Amostra composta dos elementos:

i	4
i+k	12
i+2k	20
i+3k	28
i+4k	36
i+5k	44
i+6k	52
i+7k	60
i+8k	68
i+(n-1)k	76

Se o intervalo de amostragem não for inteiro proceder da seguinte forma:

N= 321 ; n=154; $K = \frac{N}{n} = \frac{321}{154} = 2,084$; i deve ser um número sorteado entre 1 e 2,084.

Sortear um número entre 1000 e 2084 e dividir o resultado por 1000

Número sorteado = 1941, portanto i=1,941

Indivíduos:

		<u>Elementos</u>
i	1,941	1
i+k	1,941+2,084 = 4,025	4
i+2k	1,941+4,168 = 6,109	6
i+3k	1,941+6,252 = 8,193	8
.	.	.
.	.	.
.	.	.
i+(n-1)k	1,941+318,852 = 320,793	320

Amostragem casual simples estratificada com partilha proporcional

A população possui estratos com tamanhos:

$N_1; N_2; N_3$, onde a soma dos estratos é o tamanho da população, ou seja $\sum N_i = N$

A amostra deve conter os elementos da população nas mesmas proporções dos estratos. Tem-se que os tamanhos dos estratos amostrais são n_1, n_2 e n_3 tal que $\sum n_i = n$

Aplicando-se a proporção:

$$\frac{n_i}{n} = \frac{N_i}{N} \Rightarrow n_i = n \frac{N_i}{N}$$

Exemplo:

$N=500$; $N_1=50$; $N_2=150$; $N_3=300$ e $n=40$

Estrato i	Tamanho do estrato		$\frac{n_i}{n} = \frac{N_i}{N}$
	na população N_i	na amostra n_i	
1	50	4	0,1
2	150	12	0,3
3	300	24	0,6
Total	500	40	

$$n_1 = 40 \frac{50}{500} = 4; \quad n_2 = 40 \frac{150}{500} = 12; \quad n_3 = 40 \frac{300}{500} = 24$$

Amostragem por conglomerado:

O conglomerado é um conjunto de elementos formando uma unidade amostral. Se a unidade amostral for indivíduo e forem sorteados domicílios, então a amostragem é por conglomerado.

Coleta de dados, apuração de dados

Coleta de dados: é a observação e registro da categoria ou medida de variáveis relacionadas ao objeto de estudo que ocorrem em unidades (indivíduos) de uma amostra ou população.

Apuração de dados: é o processo no qual conta-se o número de vezes que a variável assumiu um determinado valor (frequência de ocorrência). Pode ser manual, mecânica ou eletrônica (programas estatísticos: Epi info, Stata, Excel, SPSS, SAS, R, S-Plus).

Distribuição de frequências - correspondência entre categorias (valores) e frequência de ocorrência.

Distribuição de frequências - correspondência entre categorias ou valores da variável e frequência de ocorrência.

Notação:

X : variável

x_i : valor observado para o indivíduo i

Exemplos de distribuição de frequências pontuais

Unidade de observação: mosquito

X: Local de captura de mosquitos (intra domicílio, peridomicílio, campo)

Distribuição de frequência dos mosquitos segundo local:

Local de captura	n
Intra domicílio	2
Peridomicílio	5
Campo	3

Exemplo

X: Número de repastos sanguíneos para completar um ciclo gonotrófico

Dados:

Mosquito	i	Valor
	1	1
	2	1
	3	2
	4	4
	5	3
	6	2
	7	3
	8	2
	9	2
	10	4

Distribuição de frequência de mosquitos segundo número de repastos:

Número de repastos	n
1	2
2	4
3	2
4	2

Banco de dados construído no pacote Jamovi utilizando o exemplo 1:

Banco de dados construído no pacote Stata utilizando o exemplo 1:

Nome da variável	Detalhamento	Códigos
id	Número de identificação do participante	
idade	Idade (anos)	
sexo	Sexo	1-masculino 2-feminino
imc	Índice de massa corporal	
dcnt	Número de doenças	
triglicerides	Concentração de triglicérides (mg/dL)	

Tela do Jamovi (banco de dados: idosos, município de São Paulo)

Distribuição de frequências com dados pontuais utilizando o Jamovi

Dados pontuais – variável qualitativa nominal e variável quantitativa discreta.

Frequencies of sexo

Levels	Counts	% of Total	Cumulative %
F	34	68.0 %	68.0 %
M	16	32.0 %	100.0 %

Variável quantitativa contínua utilizando o jamovi

Frequencies of idade

Levels	Counts	% of Total	Cumulative %
61	2	4.0 %	4.0 %
62	1	2.0 %	6.0 %
64	2	4.0 %	10.0 %
65	1	2.0 %	12.0 %
66	2	4.0 %	16.0 %
68	1	2.0 %	18.0 %
71	2	4.0 %	22.0 %
72	2	4.0 %	26.0 %
73	2	4.0 %	30.0 %
74	4	8.0 %	38.0 %
75	1	2.0 %	40.0 %
76	2	4.0 %	44.0 %
78	2	4.0 %	48.0 %
80	3	6.0 %	54.0 %
82	4	8.0 %	62.0 %
83	2	4.0 %	66.0 %
84	1	2.0 %	68.0 %
85	2	4.0 %	72.0 %
86	2	4.0 %	76.0 %
87	2	4.0 %	80.0 %
88	1	2.0 %	82.0 %
89	2	4.0 %	86.0 %
91	3	6.0 %	92.0 %
92	1	2.0 %	94.0 %
93	2	4.0 %	98.0 %
94	1	2.0 %	100.0 %

Tabelas e gráficos

- Possibilitam conhecer as características da população sob estudo porque resumem e organizam os dados.
- Permitem identificar rapidamente onde a maioria dos indivíduos está e quais são os padrões de ocorrência de valores.
- Fornecem uma idéia prévia de como serão as estimativas dos parâmetros sob investigação.
- Auxiliam na identificação dos testes estatísticos que serão efetuados em fases mais avançadas da análise dos dados.

Guia de apresentação tabular do IBGE



<http://biblioteca.ibge.gov.br/visualizacao/livros/liv23907.pdf>

Apresentação tabular

(IBGE, 1993; Berquó et al, 1981)

Elementos da tabela: título, corpo, cabeçalho, coluna indicadora, fonte e notas.

Tabela 1 - Título: o que (natureza do fato estudado)? como (variáveis)? onde? quando?

Variável	n°	%
Total		

Fonte
notas, chamadas

OBS: nenhuma casela (intersecção entre linha e coluna) deve ficar em branco.

A tabela deve ser uniforme quanto ao número de casas decimais e conter os símbolos – ou 0 quando o valor numérico é nulo e ... quando não se dispõe do dado.

Apresentação de variável qualitativa

Exemplo

Considerando-se a variável imc para classificar indivíduos segundo o estado nutricional

Tabela 1- Distribuição de idosos segundo classificação nutricional. Município de São Paulo, 2013.

Estado nutricional ⁽²⁾	n	%
Magreza	4	8,0
Eutrofia	26	52,0
Excesso de peso	20	40,0
Total	50	100

⁽²⁾ magreza: ≤ 21 kg/m²; eutrofia: 22-27 kg/m²; excesso de peso ≥ 28 kg/m²

Exemplo:

Distribuição de crianças⁽¹⁾ segundo níveis séricos de retinol. Cansação – Bahia, 1992

Nível de retinol sérico ⁽²⁾	n	%
Aceitável	89	55,3
Baixo	65	40,4
Deficiente	7	4,3
Total	161	100

⁽¹⁾ 0 –72 meses

⁽²⁾ aceitável: 20,0 – 49,9 µg/dl; baixo: 10,0 – 19,9 µg/dl; deficiente: <10,0 µg/dl

Fonte: Prado MS et al., 1995.

Exemplo:

Distribuição de Culicídeos segundo espécie coletados na área de influência indireta da Usina Hidrelétrica de Porto Primavera, SP e MS, Brasil, 1992-1993.

Táxon	n	%
<i>Culex (Culex) quinquefasciatus</i>	244	25,3
<i>Culex (Culex) sp.gr. Coronator</i>	135	14,0
<i>Culex (Culex) spp</i>	131	13,6
<i>Culex (Melanoconion) spp.</i>	111	11,5
<i>Anopheles (Nyssorhynchys) albitarsis</i>	76	7,9
<i>Culex (Culex) sp.pr. inflictus</i>	59	6,1
Outras ^(*)	210	21,7
Total	966	100

(*) Espécies ou grupos: *Aedeomyia squamipennis*, *Aedes aegypti*, *Aedes fluviatilis*, *Anopheles argyritarsis*, *Anopheles evansae*, *Anopheles oswaldoi*, *Anopheles triannulatus*, *Culex chidesteri*, *Culex camposi*, *Culex dolosus*, *Culex mollis*, *Culex saltanensis*, *Culex surinamensis*, *Culex bigoti*, *Culex sp. gr. Atratus*, *Culex aureonatus*, *Culex bastagarius*, *Culex innovator* ou *pilosus*, *Culex oedipus*, *Culex theobaldi*, *Culex vaxus*, *Psorophora albigena*, *Psorophora confinnis*, *Psorophora sp.*, *Psorophora ciliata*, *Toxorhynchites portoricensis px*, *Uranotaenia apicalis*, *Uranotaenia geometrica*, *Uranotaenia pulcherrima*, *Uranotaenia lowii*, *Uranotaenia sp.*

Fonte: Adaptado de Natal D et al., 1995. Revista brasileira de Entomologia. 39(4): 897-899.

Exercício 2

Apresentar e descrever os dados dos idosos relativos as variáveis sexo, número de doenças crônicas em tabelas de distribuição de frequência.

Variável sexo

Variável número de doenças crônicas

Apresentação tabular de variável quantitativa contínua (em intervalos de classe)

Para representar variáveis quantitativas contínuas é necessário construir intervalos de classe definidos como um conjunto de observações contidas entre dois valores limite (limite inferior e limite superior).

Os valores dos limites inferior e superior podem ou não estarem contidos no intervalo. Se um valor estiver contido a representação do intervalo deverá indicar que este é fechado naquele limite.

Por exemplo os intervalos abaixo são fechados no limite inferior. Acrescente um novo intervalo antes e após o intervalo apresentado.

5 | -- 10 Intervalo fechado no limite inferior e aberto no limite superior (contém o valor 5 mas não contém o valor 10)

Os intervalos abaixo são fechados nos limites inferior e superior. Acrescente um novo intervalo antes e após o intervalo apresentado.

5 |--| 10 Intervalo fechado nos limites inferior e superior (contém os valores e 10)

OBS: Representar o intervalo 0 |-- | 11 meses é equivalente a representá-lo como 0 |-- 12 meses.

A **amplitude do intervalo** é o tamanho do intervalo de classe.

Supor a variável idade (anos). O intervalo 5|--10 (anos) tem amplitude 5 que é igual à diferença entre os limites ($10-5=5$) e inclui as idades 5, 6, 7, 8 e 9 anos. Um indivíduo com 10,3 anos não estaria incluído neste intervalo.

A amplitude do intervalo 5|--|10 é igual a 6 porque o intervalo é fechado no 10 e inclui todos os valores 10, 1; 10,3; ...10,9; ... 10,999999 ..., mas não inclui o 11.

Idade é quantitativa contínua e, portanto, entre dois valores existem infinitos valores. Assim, não é possível saber qual é o valor que antecede o 11 (10,999999.... até o infinito). Neste caso, para o cálculo da amplitude, utiliza-se toda a informação do intervalo e por isso, seu cálculo é feito com o valor 11. A amplitude será $11-5 = 6$ (estão incluídos aí os valores 5, 6, 7, 8, 9 e 10) ou então será $(10-5)+1 = 6$.

A amplitude do intervalo e o número de intervalos dependem basicamente do problema específico e da literatura existente sobre o assunto que será utilizada para se comparar os resultados. O **ponto médio do intervalo** é calculado somando-se o limite inferior e limite superior, dividindo-se o resultado por dois.

Na construção dos intervalos de classe é necessário que eles sejam **mutuamente excludentes** (um indivíduo não pode ser classificado em dois intervalos ao mesmo tempo) e **exaustivos** (nenhum indivíduo pode ficar sem classificação).

Exemplo:

X: Peso seco de mosquitos Culicidae (mg)

Mosquito	i	Valor	i	Valor	i	Valor
	1	0,512	8	0,291	15	0,524
	2	0,670	9	0,334	16	0,389
	3	0,430	10	0,278	17	0,524
	4	0,532	11	0,227	18	0,477
	5	0,789	12	0,432	19	0,625
	6	0,459	13	0,379	20	0,532
	7	0,339	14	0,553		

Distribuição de frequência:

Peso seco (mg)	n
0,200 -- 0,300	3
0,300 -- 0,400	4
0,400 -- 0,500	4
0,500 -- 0,600	6
0,600 -- 0,700	2
0,700 -- 0,800	1

Como idade é variável quantitativa contínua, a melhor forma de apresentá-la em tabelas é utilizando intervalos de valores denominados intervalos de classe.

Ex:

x: 5, 5, 15, 20, 20, 20, 21, 21, 22, 22

idade	frequência	%
5 -- 10	2	20
10 -- 15	0	-
15 -- 20	1	10
20 -- 25	7	70
Total	10	100

X: Peso ao nascer (g)

x: 2250, 3025, 1600, 2725, 3750, 3950, 2400, 2180, 2520, 2530

Peso (g)	frequência	%
1500 --2000	1	10
2000 --2500	3	30
2500 --3000	3	30
3000 --3500	1	10
3500 --4000	2	20
Total	10	100

X: Altura de adultos (cm)

x: 1,63; 1,60; 1,59; 1,60; 1,45; 1,73; 2,05; 1,85

Altura (cm)	n	%
1,45 --1,55	1	12,5
1,55 --1,65	4	50,0
1,65 --1,75	1	12,5
1,75 --1,85	0	-
1,85 --1,95	1	12,5
1,95 --2,05	0	-
2,05 --2,15	1	12,5
Total	8	100

Ex:

X: Peso seco em (mg) de fêmeas de *Anopheles darlingi* (vetor de malária)

x: 0,10; 0,14, 0,20; 0,24; 0,26; 0,27; 0,30; 0,32; 0,34; 0,37; 0,44

Distribuição de fêmeas de *Anopheles darlingi* segundo peso seco. Sítios em Capanema, Pará, Brasil, 1995

Peso (mg)	Frequência	%
0,10 -- 0,20	2	20
0,20 -- 0,30	4	40
0,30 -- 0,40	3	30
0,40 -- 0,45	1	10
Total	10	100

Fonte: Adaptado de Lounibos, LP et. al., 1995.

Ex:

X: Tamanho do corpo de *Anopheles darlingi* = comprimento da asa em (mm)

x: 1,7; 2,2; 2,3; 2,4; 2,5; 2,6; 2,7; 2,8; 2,9; 3,0; 3,1; 3,2; 3,3; 3,4

Distribuição de fêmeas de *Anopheles darlingi* segundo tamanho do corpo. Sítios em Capanema, Pará, Brasil, 1995

Comprimento da asa (mm)	Frequência	%
1,5 -- 2,0	1	7,1
2,0 -- 2,5	3	21,4
2,5 -- 3,0	6	42,9
3,0 -- 3,5	4	28,6
Total	14	100

Fonte: Adaptado de Lounibos, LP et. al., 1995.

Exercício 4

Apresentar e descrever os dados dos idosos relativos a variável triglicérides em uma tabela de distribuição de frequência.

Tabela de dupla entrada

Distribuição de crianças⁽¹⁾ segundo níveis séricos de retinol e idade. Cansação – Bahia, 1992.

Faixa etária (meses)	Aceitável		Inadequado		Total	
	n	%	n	%	n	%
<12	5	45,5	6	54,5	11	100
12 --24	10	43,5	13	56,5	23	100
24 --36	19	54,3	16	45,7	35	100
36 --48	21	65,6	11	34,5	32	100
48 --60	16	43,2	21	56,8	37	100
60 --73	18	78,3	5	21,7	23	100
Total	89	55,3	72	44,7	161	100

⁽¹⁾ 0 –72 meses.

⁽²⁾ aceitável: 20,0 – 49,9 µg/dl; baixo: 10,0 – 19,9 µg/dl; deficiente: <10,0 µg/dl.

Fonte: Prado MS et al., 1995.

Interpretação:

Faixas etárias menores concentram mais crianças com nível inadequado de retinol sérico

Distribuição de crianças⁽¹⁾ segundo níveis séricos de retinol e idade. Cansação – Bahia, 1992.

Faixa etária (meses)	Aceitável		Inadequado		Total	
	n	%	n	%	n	%
<12	5	5,6	6	8,3	11	6,8
12 --24	10	11,2	13	18,1	23	14,3
24 --36	19	21,3	16	22,2	35	21,7
36 --48	21	23,6	11	15,3	32	19,9
48 --60	16	18,0	21	29,2	37	23,0
60 --73	18	20,2	5	6,9	23	14,3
Total	89	100	72	100	161	100

⁽¹⁾ 0 –72 meses.

⁽²⁾ aceitável: 20,0 – 49,9 µg/dl; baixo: 10,0 – 19,9 µg/dl; deficiente: <10,0 µg/dl.

Fonte: Prado MS et al., 1995.

Interpretação

Crianças com nível inadequado são mais jovens que crianças com nível adequado.

Exemplo

Distribuição de pneus com coleta de larvas de *Aedes aegypti* segundo número de larvas e predação^(*). Dar es Salaan, Tanzânia, 1973.

Número de larvas	Predador ausente		Predador presente	
	n	%	n	%
0	75	47,2	184	83,0
1 --10	51	32,1	27	12,1
11 --20	16	10,1	8	3,6
21 --50	10	6,3	3	1,3
51 --100	5	3,1	0	-
101 --300	2	1,2	0	-
Total	159	100	222	100

(*) larva de *Toxorhynchites brevialpis*

Fonte: Clementes A.N., The biology of Mosquitoes. Vol(2) pag.203, 1999.

Interpretação:

Exercício 5

Os dados a seguir são de um estudo que investiga a relação entre níveis de β -caroteno (mg/L) e hábito de fumar em gestantes.

- Calcule as frequências relativas. Fixando o 100% no total de fumantes e não fumantes.
- Calcule as frequências relativas. Fixando o 100% no total do nível de B-caroteno (MG/L).
- Interprete os resultados. Existe alguma indicação de existência de associação entre as variáveis? Justifique.

a)

Distribuição de gestantes segundo níveis de **β -caroteno (mg/L)** e hábito de fumar, Joinvile, Brasil, 2002.

β-caroteno (mg/L)	Fumante		Não Fumante		Total	
	n	%	n	%	n	%
Baixo (0 – 0,213)	46		74		120	
Normal (0,214 – 1,00)	12		58		70	
Total	58		132		190	

Fonte: Silmara Silva. Tese de Mestrado/FSP/USP.

Interpretação:

b)

Distribuição de gestantes segundo níveis de **β -caroteno (mg/L)** e hábito de fumar, Joinvile, Brasil, 2002.

β-caroteno (mg/L)	Fumante		Não Fumante		Total	
	n	%	n	%	n	%
Baixo (0 – 0,213)	46		74		120	
Normal (0,214 – 1,00)	12		58		70	
Total	58		132		190	

Fonte: Silmara Silva. Tese de Mestrado/FSP/USP.

Interpretação:

Exercício

São apresentados na tabela abaixo o local de captura de *Culex quinquefasciatus* – Intradomicílio e peridomicílio e nível socioeconômico dos habitantes de setores censitários urbanos do município de Marília, de junho de 2007 a agosto de 2008.

- a) Calcule os percentuais;
b) Interprete os dados

Nível	Intradomicílio		Peridomicílio		Total	
	n	%	n	%	n	%
Baixo	119		13		132	
Intermediário	71		0		71	
Alto	57		9		66	
Total	247		22		269	

Fonte: Telles-de-Deus, J. Hábito alimentar de *Aedes aegypti* e *Culex quinquefasciatus* e sua implicação na capacidade reprodutiva. São Paulo, 2011 [Tese de Doutorado, Faculdade de Saúde Pública da Universidade de São Paulo].