

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

A palavra estatística vem do latim *status* e significa estado. Inicialmente, era utilizada para compilar dados que descreviam características de países (Estados). Em 1662, *John Graunt* publicou estatísticas de nascimentos e mortes. A partir de então, o estudo dos eventos vitais e da ocorrência de doenças e óbitos impulsionou o desenvolvimento da Estatística nos campos teórico e aplicado (Triola, 1999).

Atualmente, índices e indicadores estatísticos fazem parte do dia a dia, tais como taxa de inflação, índice de desemprego, taxa de natalidade, taxa de crescimento populacional, índice de poluição atmosférica, índice de massa corporal, entre outros.

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

Bioestatística – Estatística aplicada às ciências da vida.

Níveis de mensuração

Escala nominal

Os elementos de análise, por exemplo indivíduos, insetos, armadilhas, são classificados em categorias segundo uma característica.

Ex:

Sexo dos insetos vetores (fêmea, macho);

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

Local do domicílio (intradomicílio, peridomicílio);

Local de dispersão espacial horizontal (mata, margem de mata, aberto, domicílio);

Local de dispersão espacial vertical (copa das árvores, solo);

Hábito alimentar (sangue de boi, sangue de aves, sangue de roedores, sangue do homem);

Tipo de criadouro (pneu, caixa d'água, vaso de planta, oco de árvore, internódio de bambu, folhigo no solo, bromélias).

A característica deste nível de mensuração é que não existe ordem entre as categorias e suas representações, se numéricas, são destituídas de significado numérico.

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*

1= mata, 2 = margem da mata, 3= campo aberto, 4= domicílio

Os valores 1, 2, 3, e 4 são apenas rótulos.

Escala ordinal

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

Ex: 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 ($\leq 2,0\text{mm}$),
Médio ($2,01 - 3,65\text{ mm}$),
Grande ($\geq 3,66\text{ mm}$).

Momento de alimentação ou atividade circadiana de *Cx. quinquefasciatus* (primeira hora, segunda hora, ...)

Embora exista ordem nas categorias, a diferença entre as categorias adjacentes não tem o mesmo significado em toda a escala.

Escala numérica intervalar

Este nível de mensuração possui 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 graus 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.

material	$^{\circ}\text{C}$	$^{\circ}\text{F}$	$ \text{dif}^{\circ}\text{C} $	$ \text{dif}^{\circ}\text{F} $	$\text{dif}^{\circ}\text{C}/\text{dif}^{\circ}\text{F}$	$\text{razão}^{\circ}\text{C}$	$\text{razão}^{\circ}\text{F}$	$\text{Razão}^{\circ}\text{C}/\text{razão}^{\circ}\text{F}$
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

comprimento	cm	polegada	$ \text{difcm} $	$ \text{dif pol} $	$\text{Difcm}/\text{difpol}$	Razãocm	Razãopol	$\text{Razãocm}/\text{razãopol}$
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

Esta escala possui zero inerente, de acordo com a natureza da característica sendo aferida. Por exemplo o comprimento da asa de um inseto. Pode ser dividida em razões contínua e razões discreta. Na escala de razões contínua o resultado numérico é um valor pertencente ao conjunto dos números reais $R = \{-\infty; \dots; 0; 0,2; 0,73; 1; 2,48; \dots; +\infty\}$.

Ex

Precipitação pluviométrica em mm (quantidade de chuva por metro quadrado)
Umidade relativa do ar (%) (razão entre o percentual em número de moléculas de água no ar pelo percentual que corresponde à saturação naquela temperatura do ambiente).
Peso seco (mg) de fêmeas ou pupas de mosquitos (0,53; 0,43;...)
Volume do repasto sanguíneo (μl) (4,7; 3,6; 4,0; 4,9 ...)

Comprimento da asa (mm)
 Largura de partes do corpo de Triatomíneos (mm)
 Tempo de sobrevivência
 Tamanho do ciclo gonotrófico

Escala de razões discreta:

O resultado numérico da mensuração é um valor inteiro.

Ex: Número de exemplares 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)

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

VARIÁVEL: $\left\{ \begin{array}{l} \text{qualitativa} \left\{ \begin{array}{l} \text{nominal} \\ \text{ordinal} \end{array} \right. \\ \text{quantitativa} \left\{ \begin{array}{l} \text{discreta} \\ \text{contínua} \end{array} \right. \end{array} \right.$

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.

Exercício 1 –

Classificar quanto a natureza, as seguintes variáveis

Variável	Tipo (natureza)
Estado de paridade da fêmea de inseto vetor (nulípara, parida)	
Tipo de abrigo (intradomiciliar, peridomiciliar)	
Taxa de paridade de fêmeas de <i>Culex quinquefasciatus</i> ⁽¹⁾	
Peso seco de pupas de <i>Aedes aegypti</i> (mg)	
Abundância numérica de flebotomíneos ao longo dos meses ⁽²⁾	
Tipo de fonte de alimentação (humano, cão, gato)	
Fases de desenvolvimento do <i>Aedes aegypti</i> (ovo, larva, pupa, adulto)	
Ocorrência de <i>Triatoma infestans</i> em determinado <i>habitat</i> (sim ou não)	
Altitude da área de coleta (m)	
Altura da copa da árvore (m)	
Classificação da altura da copa da árvore Baixa (<10m), média (10 – 29 m), alta (30 metros e mais)	
Horário de coletas (horas e minutos)	
Resultado sorológico para presença de vírus (reagente, não reagente)	
Número de larvas e pupas em determinado criadouro	
Taxa de infectividade ⁽³⁾	

$$(1) \text{ Taxa de paridade} = \frac{\text{Número de fêmeas hematófagos paridas}}{\text{Número de fêmeas paridas} + \text{nulíparas}}$$

(2) Número de indivíduos de uma determinada espécie

$$(3) \text{ Taxa de infectividade} = \frac{\text{Número de insetos infectados}}{\text{Número de insetos avaliados}}$$

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 que 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.

Elementos: são unidades de análise por exemplo pessoas, células, gen, 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 fêmeas de *A. aegypti* por picada.

Atenção: em condições ideais um mosquito ingere de 2 a 5 mg ou cerca de uma e meia a duas vezes o seu peso. Isto não é o parâmetro mas sim uma informação de um indivíduo. Para a população de *Anopheles cruzii*, por exemplo, o parâmetro seria o valor médio calculado com base nos valores individuais.

Temperatura média onde ocorre o maior número (ou taxa) de sobrevivência larval no processo de mudança de *instar* até se transformar em pupa.

Estimador: é uma fórmula matemática que permite estimar um parâmetro. Se a estimativa for um único valor, o estimador é denominado - estimador por ponto, e se a estimativa for um conjunto de valores, o estimador recebe o nome de estimador por intervalo.

Estimador por ponto:

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

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

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

Ex: X: Peso seco (mg) de fêmeas de *Culex quinquefasciatus*
 X: 0,419; 0,641; 0,592; 0,477; 0,613; 0,501
 $\bar{x} = 0,54$ mg.

Apuração de dados

A apuração de dados envolve a contagem do número de vezes que a variável assume um determinado valor determinando-se a frequência de ocorrência. A apuração pode ser manual ou eletrônica por meio da utilização de programas estatísticos tais como o Epi info, R, Stata, Excel, SPSS, SAS, e S-Plus. Os dois primeiros são de domínio público e podem ser utilizados por "down load" dos sites específicos; os demais são comerciais.

A distribuição de frequências consiste na correspondência entre categorias ou valores da variável e a frequência de ocorrência. Podem ser pontuais (variáveis qualitativas nominal e ordinal e variáveis quantitativas discreta) e em intervalos de classe (variáveis quantitativas contínua).

Notação:

X : variável

x_i : valor observado para o indivíduo i

Exemplos de distribuição de frequências pontuais

Unidade de observação: mosquito (n=10)

X: Local de captura de mosquitos (intradomicílio, peridomicílio, campo)

Dados:

Mosquito	i	x_i	categoria
	1	x_1	intradomicílio
	2	x_2	intradomicílio
	3	x_3	campo
	4	x_4	peridomicílio
	5	x_5	peridomicílio
	6	x_6	campo
	7	x_7	peridomicílio
	8	x_8	campo
	9	x_9	peridomicílio
	10	x_{10}	peridomicílio

Distribuição de frequência:

Local de captura	n
intradomicílio	2
peridomicílio	5
campo	3

Exemplo

X: Número de repastos sanguíneos para completar um ciclo gonotrófico

Dados:

Mosquito	i	x_i	Valor
	1	x_1	1
	2	x_2	1
	3	x_3	2
	4	x_4	4
	5	x_5	3
	6	x_6	2
	7	x_7	3
	8	x_8	2
	9	x_9	2
	10	x_{10}	4

Distribuição de frequência:

Número de repastos	n
1	2
2	4
3	2
4	2

Exemplos de distribuição de frequências por intervalo

Neste caso 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 estar 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 5 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). Ele inclui as idades 5, 6, 7, 8 e 9 anos mas indivíduos com 10,0 e 10,3 anos não estariam incluídos neste intervalo. A amplitude do intervalo 5|--|10 é igual a 6 porque o intervalo é fechado no 10 e inclui todos os valores até chegar no 11, mas não inclui o 11. A variável sendo

contínua indica que entre dois valores existem infinitos valores então não é possível saber qual é o valor que antecede o 11 (10,999999.... até o infinito). Neste caso a amplitude utiliza 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 $(10-5)+1=6$.

A amplitude do intervalo e o número de intervalos dependem basicamente do problema específico e da literatura existente sobre o assunto.

Na construção dos intervalos de classe é necessário que eles sejam **mutuamente exclusivos** (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	xi	Valor	i	xi	Valor	i	xi	Valor
	1	x_1	0,512	8	x_8	0,291	15	x_{15}	0,524
	2	x_2	0,670	9	x_9	0,334	16	x_{16}	0,389
	3	x_3	0,430	10	x_{10}	0,278	17	x_{17}	0,524
	4	x_4	0,532	11	x_{11}	0,227	18	x_{18}	0,477
	5	x_5	0,789	12	x_{12}	0,432	19	x_{19}	0,625
	6	x_6	0,459	13	x_{13}	0,379	20	x_{20}	0,532
	7	x_7	0,339	14	x_{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

Exercício 2

Apure os dados abaixo apresentando a variável em intervalos de classe. Para construir a distribuição de frequência, conte quantos insetos caem em cada classe.

X: Comprimento da asa de mosquitos Culicidae (mm)

2,30	2,34	2,40	2,47	2,54	2,60	2,64	2,64	2,68	2,71
2,74	2,74	2,74	2,74	2,74	2,76	2,76	2,76	2,79	2,79
2,82	2,83	2,84	2,84	2,84	2,89	2,90	2,92	2,93	2,93
2,93	2,93	2,93	2,94	2,94	2,94	2,94	2,94	2,94	2,94
2,99	3,00	3,02	3,02	3,02	3,02	3,02	3,02	3,04	3,04
3,04	3,04	3,04	3,04	3,04	3,04	3,04	3,04	3,05	3,05
3,05	3,05	3,05	3,08	3,08	3,09	3,09	3,09	3,09	3,11
3,11	3,11	3,11	3,11	3,12	3,12	3,12	3,12	3,14	3,14
3,14	3,14	3,15	3,15	3,17	3,19	3,19	3,20	3,20	3,20
3,20	3,24	3,24	3,24	3,24	3,29	3,29	3,29	3,29	3,31
3,31	3,34	3,39	3,44						

Apresentação tabular de variável quantitativa contínua

Como a variável peso seco (mg) é quantitativa contínua, a representação tabular apropriada é em intervalos de classe

Exemplo:

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

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 seco (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.

Exemplo:

X: tamanho do corpo de *Anopheles darlingi*.

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 3

Utilize a distribuição de frequências construída no exercício 2 e apresente-a em uma tabela. Os dados são de fêmeas de *Aedes vigilax* mantidas em laboratório na Escola de Biologia da Universidade de Queensland, Brisbane-Austrália, 2008. O exercício foi adaptado a partir de artigo publicado no Journal of Medical Entomology, vol.45, nº 3, 353-359 que teve como autores: Hugo L.E et. al., 2008.

- Represente os dados numa tabela
- Interprete os resultados.
- Supondo que não fossem conhecidos os tamanhos de 15 espécimes. Como você representaria esses valores?
- Ao apresentar os dados em uma tabela você iria incluir estes 15 espécimes?
- Estas 15 fêmeas tinham, na verdade, comprimento de asa maior que 3,50 mm e o investigador, por achar que eram "valores esquisitos" resolveu excluí-los. Você concorda com esta decisão? Justifique.

Tabela de dupla entrada

Exemplo

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

Número de larvas	Predador ausente		Predador presente		Total	
	n	%	n	%	n	%
0	75	47,2	184	83,0	259	68,0
1 -- 10	51	32,1	27	12,1	78	20,5
11 -- 20	16	10,1	8	3,6	24	6,3
21 -- 50	10	6,3	3	1,3	13	3,4
51 -- 100	5	3,1	0	-	5	1,3
101 -- 300	2	1,2	0	-	2	0,5
Total	159	100	222	100	381	100

(*) larva de *Toxorhynchites brevipalpis*

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

Exemplo

Distribuição de espécimes de *Aedes scapularis* segundo o tamanho do corpo e sexo. Pariquera-Açu, São Paulo, 2011

Tamanho do centróide em (mm)	Machos		Fêmeas		Total	
	n	%	n	%	n	%
1,3 --1,6	5	45,5	6	54,5	11	100
1,6 --1,9	10	43,5	13	56,5	23	100
1,9 --2,2	19	54,3	16	45,7	35	100
2,2 --2,5	21	65,6	11	34,5	32	100
2,5 --2,8	16	43,2	21	56,8	37	100
2,8 --3,1	9	64,3	5	35,7	14	100
Total	80	52,6	72	47,4	152	100

Fonte: Devicari et al. 2013 (dados adaptados)

Tamanho do centróide em (mm)	Machos		Fêmeas		Total	
	n	%	n	%	n	%
1,3 --1,6	5	6,3	6	8,3	11	7,3
1,6 --1,9	10	12,5	13	18,1	23	15,1
1,9 --2,2	19	23,8	16	22,2	35	23,0
2,2 --2,5	21	26,2	11	15,3	32	21,1
2,5 --2,8	16	20,0	21	29,2	37	24,3
2,8 --3,1	9	11,2	5	6,9	14	9,2
Total	80	100	72	100	152	100

Exercício 4

Os dados a seguir são de um estudo que investiga o tamanho do corpo e o estado de paridade de fêmeas de *Aedes triseriatus* coletados em setembro de 1985.

- Calcule as frequências relativas. Fixando o 100% no total do tamanho do corpo (mm).
- Calcule as frequências relativas. Fixando o 100% no total de fêmeas paridas e não paridas (nulíparas).
- Interprete os resultados. Existe alguma indicação de existência de associação entre as variáveis? Justifique.

Distribuição de fêmeas de *Ae. triseriatus* segundo tamanho do corpo e taxa de paridade. Wisconsin, Estados Unidos, 1985.

Tamanho do corpo (mm) (X)	Paridas		Não paridas		Total	
	n	%	n	%	n	%
Pequeno ($X \leq 2,90$)	23		25		48	
Médio ($2,91 \leq X \leq 3,65$)	127		113		240	
Grande ($X \geq 3,66$)	12		28		40	
Total	162		166		328	

Fonte: Landry SV, et.al., 1988. (Adaptado). Journal of the American Mosquito Control Association vol4, nº2.

Distribuição de fêmeas de *Ae. triseriatus* segundo tamanho do corpo e taxa de paridade. Wisconsin, Estados Unidos, 1985.

Tamanho do corpo (mm) (X)	Paridas		Não paridas		Total	
	n	%	n	%	n	%
Pequeno ($X < 2,90$)	23		25		48	
Médio ($2,90 \leq X \leq 3,65$)	127		113		240	
Grande ($X \geq 3,66$)	12		28		40	
Total	162		166		328	

Fonte: Landry SV, et.al., 1988. (Adaptado). Journal of the American Mosquito Control Association vol4, nº2.

Exercício 5

Os dados a seguir são adaptados do estudo realizado em 2003 no Município de Pedrinhas, Vale do Ribeira, estado de São Paulo e referem-se ao número de repastos sanguíneos realizados por 50 fêmeas de *Aedes albopictus*, procedentes de larvas coletadas naquela localidade, em condições de laboratório após a primeira oviposição.

Número de repastos:

2	3	2	1	2	5	5	4	3
1	2	2	1	2	5	5	4	3
2	2	3	2	3	4	2	3	2
3	2	3	3	3	4	3	4	5
3	1	4	3	4	4	3		
3	1	5	4	4	2	4		

Fonte: Fernandez, Z e Forattini, OP. 2003 (Adaptado). Revista de Saúde Pública 2003; 37(3):285-91.

- Apresente os dados em uma tabela;
- Interprete a tabela.

Exercício 6

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.

- Calcule os percentuais;
- Interprete os dados

Nível socioeconômico	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].

Exercício 7

Os dados são adaptados de artigo publicado por Devicari et al. 2013 e se referem ao tamanho das asas (mm) de *Aedes scapularis* para machos e fêmeas capturados no município de São Paulo em 2011.

Machos

1,78	1,91	2,02	2,11	2,13	2,21	2,3	2,41
1,87	2,01	2,03	2,11	2,15	2,21	2,31	3,50
1,90	2,01	2,10	2,11	2,15	2,21	2,32	

Fêmeas

1,01	1,62	2,30	2,40	2,56	2,61	2,71	2,80
1,52	1,89	2,31	2,45	2,60	2,65	2,75	2,89
1,58	1,97	2,34	2,45	2,60	2,70	2,78	

- Quais variáveis estão sendo estudadas? Identifique a natureza de cada variável;
- Apure os dados e apresente a variável tamanho das asas, em intervalos de classe, em uma tabela contendo as duas variáveis;
- Classifique a variável tamanho das asas em duas categorias: pequenas (1,00 a 1,90 mm), médias (1,91 a 2,30 mm) e grande (2,31 a 3,60 mm) e faça uma tabela bidimensional cruzando as variáveis. Interprete os resultados.

Exercício 8

Os dados a seguir são relativos ao peso seco (mg) de fêmeas de *Culex quinquefasciatus* cujas larvas foram tratadas com mistura de ração de peixe, leite ninho e ração de cão, submetidas a temperaturas de 20°C (*) e acima de 20°C.

0,62*	0,77*	0,84*		0,48	0,61	0,64	0,72
0,64*	0,77*	0,94*		0,50	0,61	0,64	0,72
0,65*	0,79*			0,50	0,62	0,64	0,72
0,70*	0,79*			0,59	0,62	0,66	0,73
0,72*	0,80*			0,59	0,62	0,66	0,73
0,73*	0,80*			0,59	0,62	0,70	0,73
0,73*	0,81*			0,60	0,62	0,70	0,74
0,74*	0,83*			0,61	0,64	0,70	0,75

Fonte: Marchi MJ. Padronização de técnica para produção em massa de *Culex quinquefasciatus* (Diptera: Culicidae). São Paulo, 2014 [Dissertação de Mestrado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

(*) = larvas submetidas a temperatura de 20°C.

- Classifique a variável peso seco (mg) em duas categorias: baixo peso (abaixo de 0,65 mg) e peso não baixo (0,65 mg e mais) e faça uma tabela bidimensional cruzando com a variável temperatura (igual a 20°C e acima de 20°C).
- Interprete os resultados.

Exercício 9

A tabela abaixo foi extraída de estudo que objetivou estudar pacientes com suspeita de dengue ou febre amarela de seis mesorregiões do Estado do Pará.

Tabela 1 - Distribuição da positividade do teste de inibição da hemaglutinação para Flavivirus, por mesorregião, Pará, jun/dez 1999.

Mesorregião	Positivos		Negativos		Total
	n°	%	n°	%	
Baixo Amazonas	15	51,7	14	48,3	29
Marajó	8	88,9	1	11,1	9
Metropolitana	203	79,0	54	21,1	257
Nordeste	77	61,1	49	38,9	126
Sudeste	248	71,5	99	28,5	347
Sudoeste	12	70,6	5	29,4	17
Total	563	71,7	222	28,3	785

Fonte: Fichas epidemiológicas – Seção de Arbovírus/IEC, junho a dezembro de 1999

Fonte: Araújo TP et al., 2002. Revista Brasileira de Medicina Tropical 35(6):579-584, 2002.

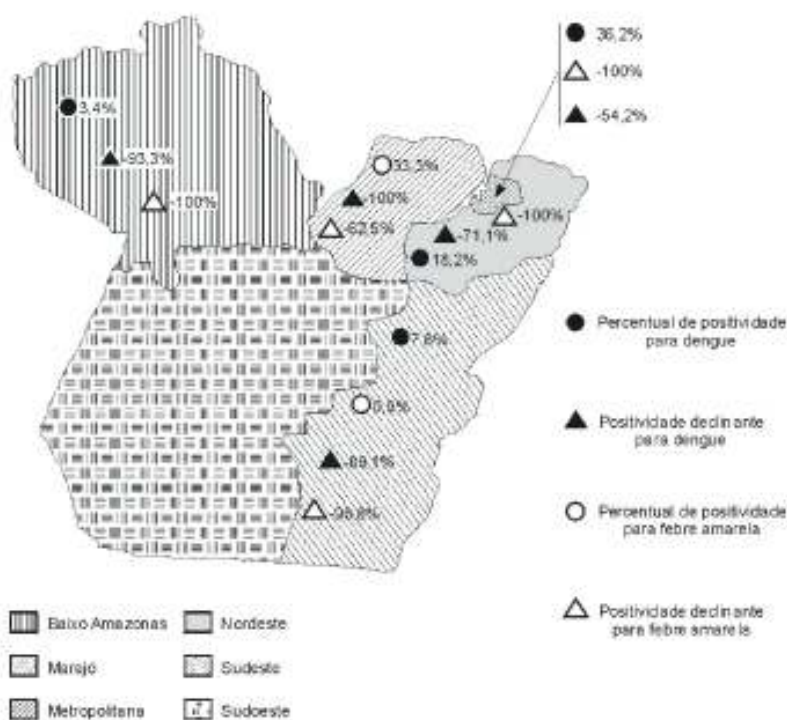


Figura 1 - Distribuição dos percentuais de positividade, em relação ao total de sôros, e declínio da positividade observada entre os resultados de IH e ELISA-IgM, por mesorregião, Pará, junho a dezembro 1999.

Fonte: Fichas epidemiológicas – Seção de Arbovírus/IEC, junho a dezembro de 1999.

Exercício 10

A tabela abaixo foi extraída de artigo que estuda aspectos da ecologia de flebotomíneos capturados em um foco de Leishmaniose tegumentar no município de Varzelândia no Estado de Minas Gerais. Comente os resultados apresentados.

Tabela 1 - Flebotomíneos capturados com armadilha tipo CDC no Brejo do Mutambal, município de Varzelândia, segundo sexo e ambiente, janeiro a dezembro de 2000.

Intradomicílio	Peridomicílio		Porcentagem		Total	
Espécies	♂	♀	♂	♀	nº	%
<i>Brumptomyia avellari</i>	1	0	1	0	2	0,1
<i>Lutzomyia capixaba</i>	0	0	1	3	4	0,1
<i>L. cavernicola</i>	3	2	108	234	347	5,1
<i>L. evandroi</i>	2	3	16	2	23	0,3
<i>L. intermedia</i>	95	20	158	70	343	5,1
<i>L. ischnacantha</i>	9	1	218	49	277	4,1
<i>L. lenti</i>	10	5	69	59	143	2,1
<i>L. longipalpis</i>	81	51	1.813	412	2.357	34,8
<i>L. migonei</i>	1	0	22	6	29	0,4
<i>L. minasensis</i>	2	4	10	13	29	0,4
<i>L. peresi</i>	13	13	25	17	68	1,0
<i>L. quinquefer</i>	15	19	71	73	178	2,6
<i>L. renei</i>	54	86	941	1.242	2.323	34,3
<i>L. sallesi</i>	4	13	27	69	113	1,7
<i>L. serrana</i>	0	0	2	1	3	0,1
<i>L. shannoni</i>	1	0	5	0	6	0,1
<i>L. sordellii</i>	0	0	1	1	2	0,1
<i>L. trinidadensis</i>	3	0	4	0	7	0,1
<i>L. whitmani</i>	2	0	2	0	4	0,1
<i>Lutzomyia spp</i>	12	36	92	358	498	7,4
Total	308	253	3.586	2.609	6.756	100,0

Fonte: Dias ES et. al., 2007. Revista da Sociedade Brasileira de Medicina Tropical 40(1)- 49-52, 2007.

Apresentação gráfica: diagrama de barras, diagramas de setores circulares, diagrama linear, histograma, polígono de frequência, ogiva de frequências acumuladas.

Diagrama de barras

Utilizado para representar variáveis qualitativa nominal, ordinal e quantitativa discreta.

Características do diagrama: é construído com figuras geométricas (barras) separadas e bases de mesmo tamanho. A altura das barras é proporcional às frequências.

Exemplos

Diagrama de barras representando uma variável qualitativa nominal

Exemplo 1

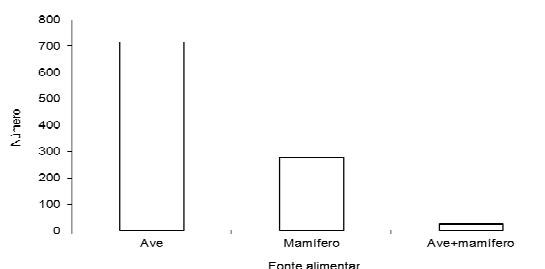
Estudo que objetivou detectar o sangue ingerido por fêmeas de mosquitos Culicidae, principalmente das fêmeas da espécie *Culex pipiens*, em área suburbana de Chicago, Illinois de 2005 a 2007. Qual é a preferência alimentar dos mosquitos desta família?

Número e percentual de fêmeas de mosquitos Culicidae, segundo fonte alimentar, coletados em área suburbana do sudoeste de Chicago em Illinois de 2005 a 2007.

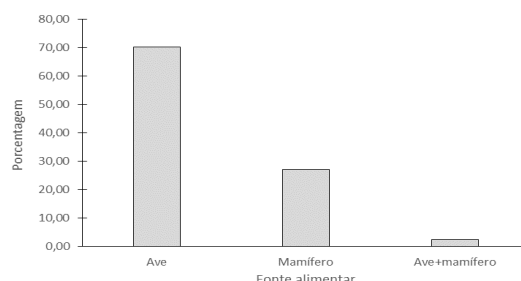
Fonte alimentar	n	%
Sangue de ave	715	70,3
Sangue de mamífero	277	27,2
Sangue de ave+mamífero*	25	2,5
Total	995	100

*repasto misto

Fonte: adaptado de Hamer GL et al., 2009. Am. Mosq.Trop.Med.Hyg., 80(2), 2009, PP.268-278.



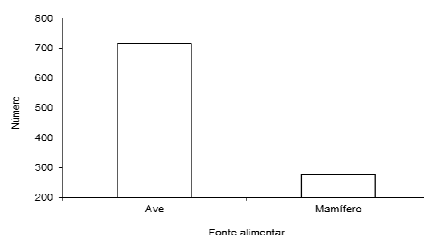
Ou



Fonte: adaptado de Hamer GL et al., 2009. Am. Mosq.Trop.Med.Hyg., 80(2), 2009, PP.268-278.

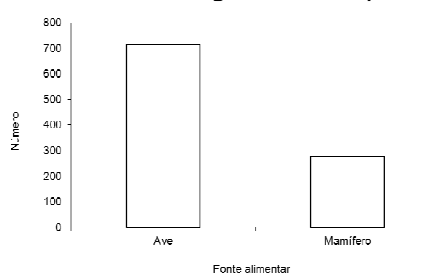
Número e percentual de fêmeas de mosquitos Culicidae, segundo fonte alimentar, coletados em área suburbana do sudoeste de Chicago em Illinois de 2005 a 2007.

Se fosse de interesse representar somente as duas primeiras categorias, esta representação gráfica está correta?



Atenção: cuidado com a origem!

Diagrama de barras da tabela anterior, excluindo-se os registros da categoria sem o repasto misto de ave+mamífero e adotando-se o valor zero na origem do eixo y.



Fonte: adaptado de Hamer GL et al., 2009. Am. Mosq.Trop.Med.Hyg., 80(2), 2009, PP.268-278.

Número e percentual de fêmeas de mosquitos Culicidae, segundo fonte alimentar, coletados em área suburbana do sudoeste de Chicago em Illinois de 2005 a 2007.

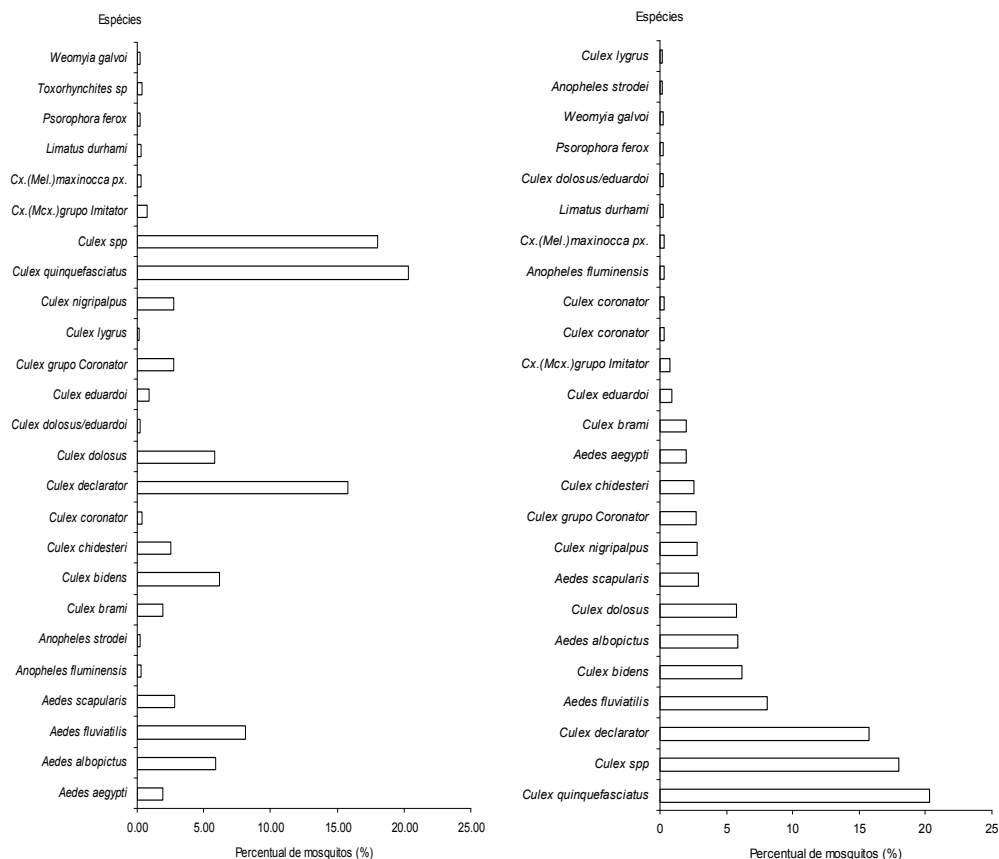
Exemplo 2

Distribuição do percentual de mosquitos Culicidae segundo espécie, coletados em 35 parques municipais de São Paulo, no período de outubro de 2010 a fevereiro de 2011.

Categoria Taxonômica	Percentual*	Categoria Taxonômica	Percentual*
<i>Culex quinquefasciatus</i>	20,28	<i>Culex eduardoi</i>	0,92
<i>Culex spp</i>	17,98	Cx.(Mcx.)grupo Imitator	0,74
<i>Culex declarator</i>	15,75	<i>Culex coronator</i>	0,37
<i>Aedes fluviatilis</i>	8,09	<i>Toxorhynchites spp</i>	0,37
<i>Culex bidens</i>	6,16	<i>Anopheles fluminensis</i>	0,31
<i>Aedes albopictus</i>	5,85	Cx.(Mel.)maxinocca px.	0,29
<i>Culex dolosus</i>	5,77	<i>Limatus durhami</i>	0,27
<i>Aedes scapularis</i>	2,85	<i>Psorophora ferox</i>	0,21
<i>Culex nigripalpus</i>	2,77	<i>Culex dolosus/eduardoi</i>	0,21
Culex grupo Coronator	2,73	<i>Weomyia galvoi</i>	0,21
<i>Culex chidesteri</i>	2,55	<i>Anopheles strodei</i>	0,19
<i>Aedes aegypti</i>	1,95	<i>Culex lygrus</i>	0,18
<i>Culex brami</i>	1,95		

* percentual em relação ao total de mosquitos imaturos e adultos coletados.

Fonte: Medeiros-Souza, AR et al., 2011 (adaptado). Biota Neotrop., vol. 13, no. 1, 317-321.



Fonte: Medeiros-Souza, AR et al., 2011 (adaptado). Biota Neotrop., vol. 13, no. 1, 317-321.

Distribuição do percentual de mosquitos imaturos e adultos (Diptera:Culicidae) coletados em 35 parques municipais da cidade de São Paulo.

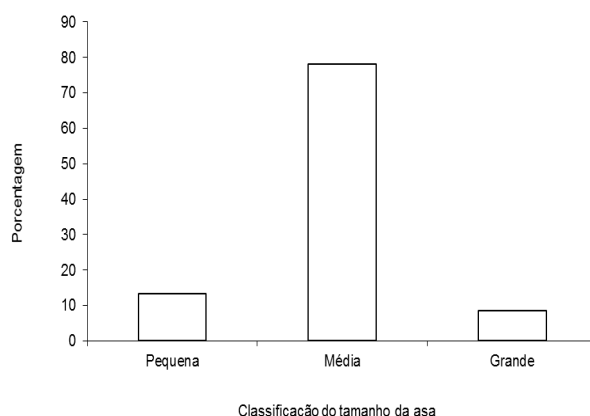
Diagrama de barras representando uma variável qualitativa ordinal

Distribuição do número de fêmeas de *Aedes triseriatus* segundo classificação do comprimento da asa (mm), Iowa Co., Wisconsin, Madison em 1988.

Classificação do comprimento da asa (mm)	n	%
Pequena	102	13,3
Média	601	78,1
Grande	66	8,6
Total	769	100

Fonte: Adaptado de Landry SV et al., 1988

Journal of the American Mosquito Control Association, 1988, Vol4, nº 2, 121-128.



Fonte: Adaptado de Landry SV et al., 1988. Journal of the American Mosquito Control Association, 1988, Vol4, nº 2, 121-128
Distribuição do número de fêmeas de *Aedes triseriatus* segundo classificação do comprimento da asa (mm), Iowa Co., Wisconsin, Madison em 1988.

Exercício 11

Como você descreveria fêmeas desta espécie segundo o comprimento das asas?

Diagrama de barras para representar uma variável quantitativa discreta:

Foi realizado experimento para determinação do padrão temporal diário de oviposição de fêmeas de *Aedes aegypti* em laboratório. As fêmeas não exibiram nenhuma atividade de oviposição nas 48 horas subseqüentes à primeira alimentação sanguínea. Apenas ocorreu oviposição no terceiro dia após o primeiro repasto sanguíneo sendo este o de maior percentual em relação aos dias subseqüentes.

O que está sendo estudado: número de ovos postos

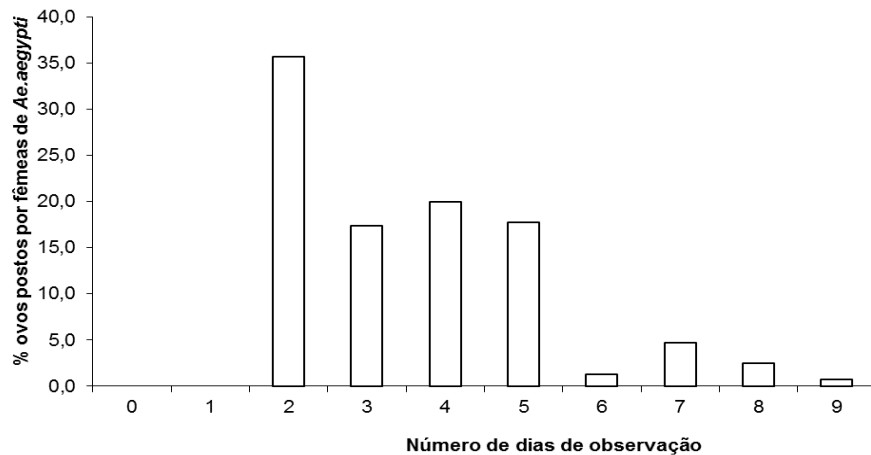
Variável: Número de dias de observação

Distribuição de ovos postos por fêmeas de *Aedes aegypti* segundo número de dias de observação após o primeiro repasto sanguíneo. Campus Pampulha da Universidade federal de Minas Gerais. Março, 2002.

Número de dias de observação*	n	%
0	0	-
1	0	-
2	2746	35,7
3	1341	17,4
4	1530	20,0
5	1362	17,7
6	96	1,3
7	363	4,7
8	191	2,5
9	56	0,7
Total	7685	100

* após a primeira alimentação das fêmeas de *Aedes aegypti* com sangue.

Fonte: (adaptado) Gomes AS, Sciavico J C S. Eiras, AE. Revista da Sociedade Brasileira de Medicina Tropical 39(4):327-332, julho, 2006



Distribuição de ovos postos por fêmeas de *Aedes aegypti* segundo dias de observação após o primeiro repasto sanguíneo. Campus Pampulha da Universidade federal de Minas Gerais. Março, 2002.

Exercício 12

Os dados a seguir são relativos ao número de repastos sanguíneos realizados por 50 fêmeas de *Aedes albopictus*.

2	3	2	1	2	5	5	4	3
1	2	2	1	2	5	5	4	3
2	2	3	2	3	4	2	3	2
3	2	3	3	3	4	3	4	5
3	1	4	3	4	4	3		
3	1	5	4	4	2	4		

- Apresente os dados em um gráfico.
- Interprete o gráfico.

Diagrama de setores circulares

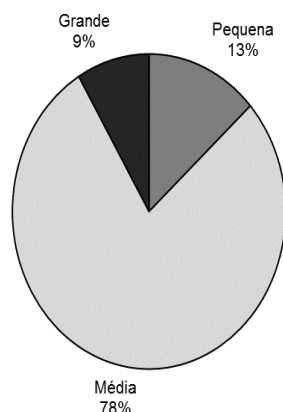
Variáveis: qualitativa nominal e qualitativa ordinal

Distribuição do número de fêmeas de *Aedes triseriatus* segundo classificação do comprimento da asa (mm). Wisconsin, Madison em 1988.

Classificação do comprimento da asa (mm)*	n	%
Pequena	102	13,3
Média	601	78,1
Grande	66	8,6
Total	769	100

* equivalente ao tamanho do corpo do mosquito.

Fonte: Adaptado de Landry SV et al., 1988. Journal of the American Mosquito Control Association, 1988, Vol4, nº 2, 121-128.



Fonte: Adaptado de Landry SV et al., 1988. Journal of the American Mosquito Control Association, 1988, Vol4, nº 2, 121-128.

Distribuição do número de fêmeas de *Aedes triseriatus* segundo classificação do comprimento da asa coletadas in Iowa Co., Wisconsin, Madison em 1988.

Diagrama linear

Representa variáveis qualitativas ordinais com natureza contínua subjacente às categorias. Por exemplo, a variável dia da semana. As categorias segunda-feira, terça-feira, etc são rótulos (nomes) dados para cada dia da semana caracterizando uma variável qualitativa ordinal e portanto poderia ser representada por um diagrama de barras. Entretanto, por existir, de modo subjacente uma continuidade entre as categorias (quando termina a segunda-feira, imediatamente começa a terça-feira), esta variável constitui uma exceção na representação das qualitativas podendo-se unir os pontos resultando em uma linha de tendência.

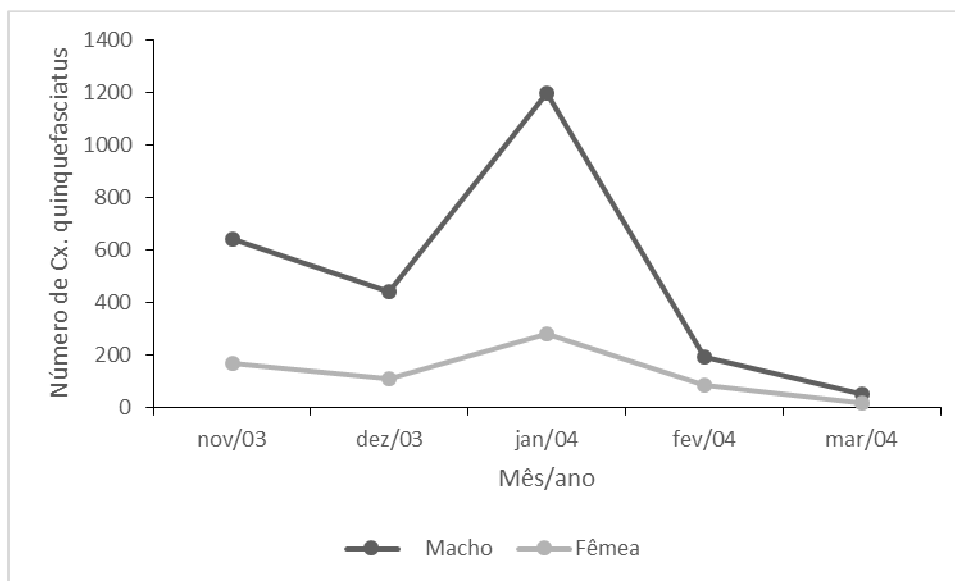
Distribuição mensal do número de *Culex quinquefasciatus* segundo sexo. São Paulo, novembro de 2003 a março de 2004.

Mês/Ano	Macho	Fêmea
Nov/2003	644	168
Dez/2003	443	112
Jan/2004	1198	284
Fev/2004	192	87
Mar/2004	53	21

Fonte: Adaptado de Laporta et.al.2006. Revista Brasileira de Entomologia 50(1):125-127, março 2006.

Exercício 13

Quais são as variáveis que estão sendo representadas?



Distribuição mensal do número de *Culex. quinquefasciatus* segundo sexo. São Paulo, novembro de 2003 a março de 2004.

Fonte: Adaptado de Laporta et.al. 2006.Revista Brasileira de Entomologia 50(1):125-127, março 2006.

Exercício 14

Os dados são referentes a taxa de incidência de dengue (número de casos confirmados, por 100 mil habitantes) nos Estados de São Paulo e Rio de Janeiro, Região Sudeste do Brasil. Período de 2004 a 2014

Ano	São Paulo	Rio de Janeiro
2004	7,8	8,2
2005	14,3	8,9
2006	130,8	171,0
2007	221,6	367,7
2008	17,9	1249,1
2009	21,8	42,5
2010	503,0	186,5
2011	278,4	1036,8
2012	69,6	1116,2
2013	506,0	1301,6
2014	515,2	46,9

Fonte: SES/SINAN

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

- Apresente os dados em um gráfico.
- Interprete os resultados.

Gráfico polar

Apropriado para representar variáveis qualitativas cujas categorias apresentam padrão de repetição como no caso dos dias da semana, estações do ano, período crepuscular (matutino e vespertino). Permite identificar padrões na representação das categorias.

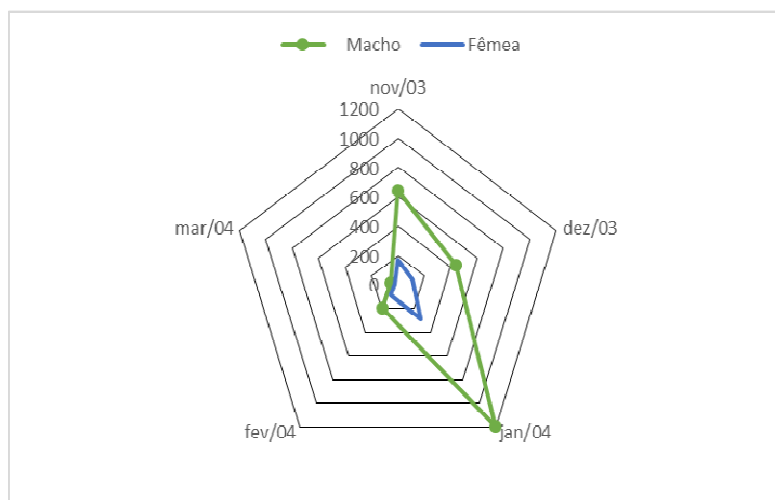
O gráfico polar representa os dados em raios em número igual ao das categorias. A origem da escala de cada raio é no centro do círculo. Os pontos são unidos nas frequências de cada categoria.

Exemplo

Distribuição mensal do número de *Culex quinquefasciatus* segundo sexo. São Paulo, novembro de 2003 a março de 2004.

Mês/Ano	Macho	Fêmea
Nov/2003	644	168
Dez/2003	443	112
Jan/2004	1198	284
Fev/2004	192	87
Mar/2004	53	21

Fonte: Adaptado de Laporta et.al.2006. Revista Brasileira de Entomologia 50(1):125-127, março 2006.



Fonte: Adaptado de Laporta et.al.2006. Revista Brasileira de Entomologia 50(1):125-127, março 2006.

Distribuição mensal do número de *Culex quinquefasciatus* segundo sexo. São Paulo, novembro de 2003 a março de 2004.

Exercício 15

Representar os dados do exercício 14 utilizando o gráfico polar.

Histograma

Adequado para representar variável quantitativa contínua

Este gráfico é construído com barras justapostas com alturas proporcionais à frequência de ocorrência dos valores da variável em cada intervalo de classe sendo que a altura das barras depende da amplitude dos intervalos de classe. Por ser construído com figuras geométricas, o que se observa são as áreas ou superfície das barras. Como a área de um retângulo é dada pela altura x base, se as bases forem iguais (mesma amplitude de classe), a interpretação será guiada pela altura das barras. Se estas forem de tamanhos diferentes, não será mais possível olhar para as alturas, sendo necessário um ajuste para amplitudes diferentes. Este ajuste é dado pelo número de observações no intervalo de classe dividido pela amplitude, denominado densidade. Se este valor ajustado for menor do que um é possível multiplicá-lo por 10, 100, 1000, etc. sem alterar a interpretação. O ajuste representa o número de unidades observadas por unidade de medida. Por exemplo, na aferição do

tamanho da asa, se ocorrerem 10 insetos na classe de 1,50 a 2,99, a amplitude seria igual a 1,50 e o ajuste seria igual a 0,15 insetos por mm ou $0,15 \times 100 = 15$ insetos por mm por 100.

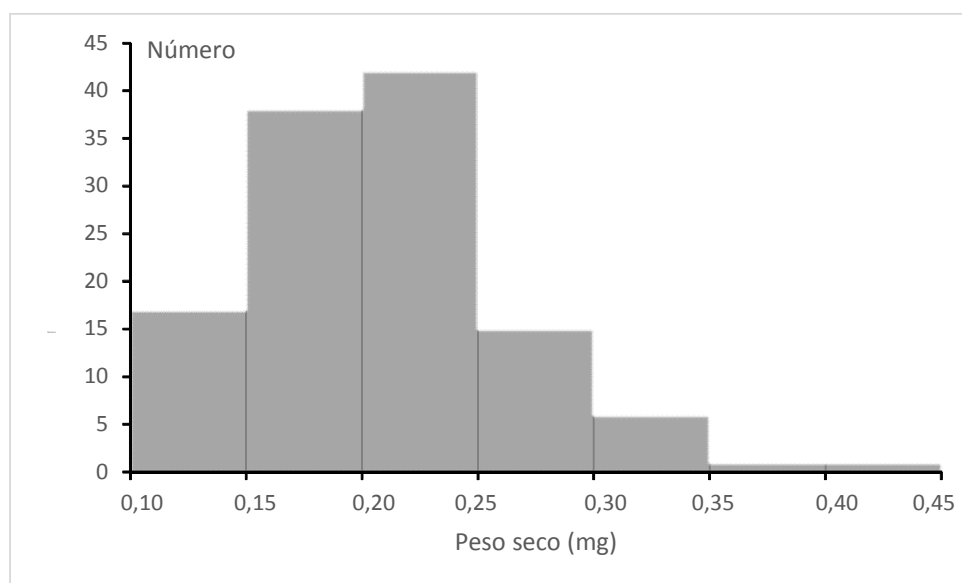
Intervalos de classe com mesma amplitude

Distribuição de fêmeas de *Anopheles darlingi*, criadas em laboratório, segundo peso seco (mg). Capanema, Estado do Pará, Brasil. 1995

Classes de peso seco (mg)	n	%
0,10 -- 0,15	17	14,2
0,15 -- 0,20	38	31,7
0,20 -- 0,25	42	35,0
0,25 -- 0,30	15	12,5
0,30 -- 0,35	6	5,0
0,35 -- 0,40	1	0,8
0,40 -- 0,45	1	0,8
Total	120	100

Fonte: Adaptado de Lounibos LP et al., 1995

1 kg = 1000 gramas; 1 g = 1000 mg

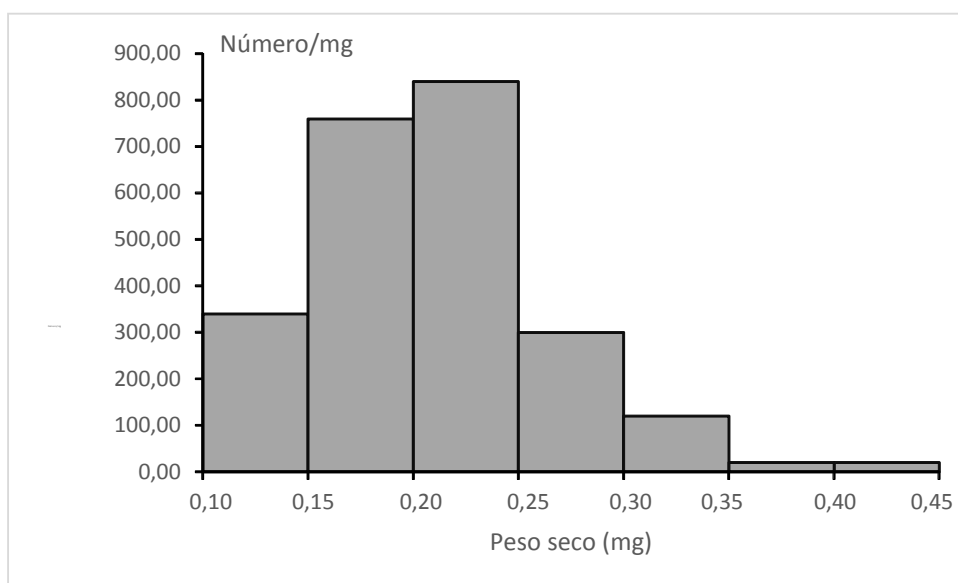


Fonte: Adaptado de Lounibos LP et al., 1995

Distribuição de fêmeas de *Anopheles darlingi*, criadas em laboratório, segundo peso seco (mg). Capanema, Estado do Pará, Brasil. 1995

Notar que o gráfico pode ser construído considerando-se o número de mosquitos por unidade de medida (densidade), no caso, mg

Classes de peso seco (mg)	n	amplitude	ajuste
0,10 -- 0,15	17	0,05	340,00
0,15 -- 0,20	38	0,05	760,00
0,20 -- 0,25	42	0,05	840,00
0,25 -- 0,30	15	0,05	300,00
0,30 -- 0,35	6	0,05	120,00
0,35 -- 0,40	1	0,05	20,00
0,40 -- 0,45	1	0,05	20,00
Total	120		



Fonte: Adaptado de Lounibos LP et al., 1995

Distribuição de fêmeas de *Anopheles darlingi*, criadas em laboratório, segundo peso seco (mg). Capanema, Estado do Pará, Brasil. 1995

Exercício 16

Distribuição de fêmeas de *Coquillettidia venezuelensis* segundo taxa de paridade. Paramaribu, Suriname, 1975

Taxa de paridade (por 100)	Nº	%
3,0 - 9,0	493	15,1
9,0 - 15,0	1193	36,5
15,0 - 21,0	1293	39,5
21,0 - 27,0	36	1,1
27,0 - 33,0	194	5,9
33,0 - 39,0	63	1,9
Total	3248	100

Fonte: Panday RS, 1975. Mosquito News vol.35 nº 3, 1975.

$$\text{Taxa de paridade} = \frac{\text{Número de fêmeas hematófagos paridas}}{\text{Número de fêmeas paridas + nulíparas}}$$

- Apresente os dados em um histograma.
- Interprete os resultados.

Intervalos de classe com amplitudes diferentes

Distribuição de fêmeas de *Aedes triseriatus* segundo comprimento da asa em (mm). Iowa Co. Wisconsin, 1988.

Comprimento da asa (mm)	Nº	%
2,50 --2,75	195	6,4
2,75 --3,15	701	22,9
3,15 --3,30	240	7,8
3,30 --3,50	1371	44,8
3,50 --3,65	381	12,4
3,65 --3,75	176	5,7
Total	3064	100

Fonte: Adaptado de Landry SV et al. 1988. Journal of the American Mosquito Control Association, 4(2),1988.

Ajuste

Comprimento da asa (mm)	Nº	Amplitude	Nº/amplitude
2,50 --2,75	195	0,25	780,0
2,75 --3,15	701	0,40	1752,5
3,15 --3,30	240	0,15	1600,0
3,30 --3,50	1371	0,20	6855,0
3,50 --3,65	381	0,15	2540,0
3,65 --3,75	176	0,10	1760,0
Total	3064		

Fonte: Adaptado de Landry SV et al. 1988. Journal of the American Mosquito Control Association, 4(2),1988.

Distribuição de fêmeas de *Aedes triseriatus* segundo comprimento da asa em (mm). Iowa Co. Wisconsin, 1988

Cuidado: Sem fazer o ajuste, o gráfico fica errado e pode levar a conclusões incorretas.

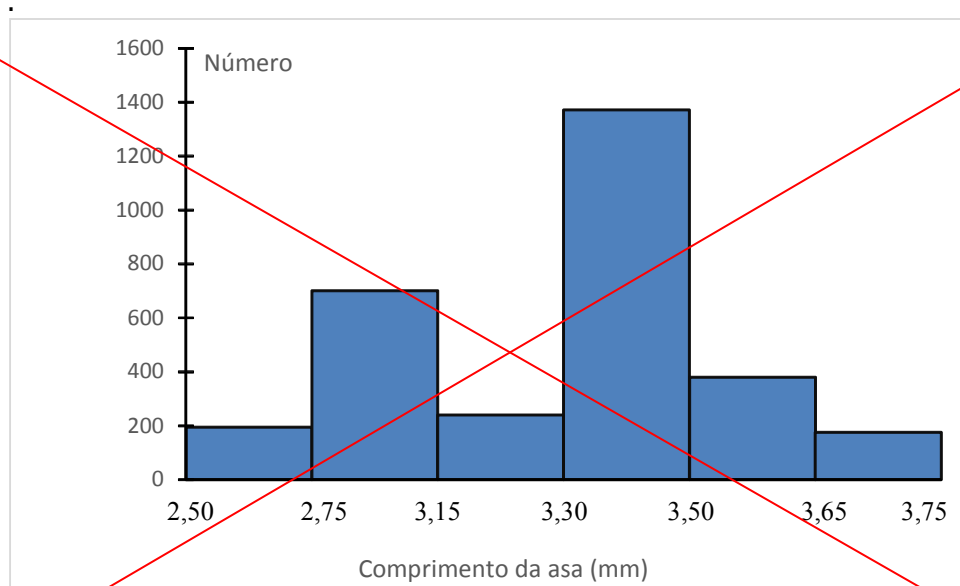
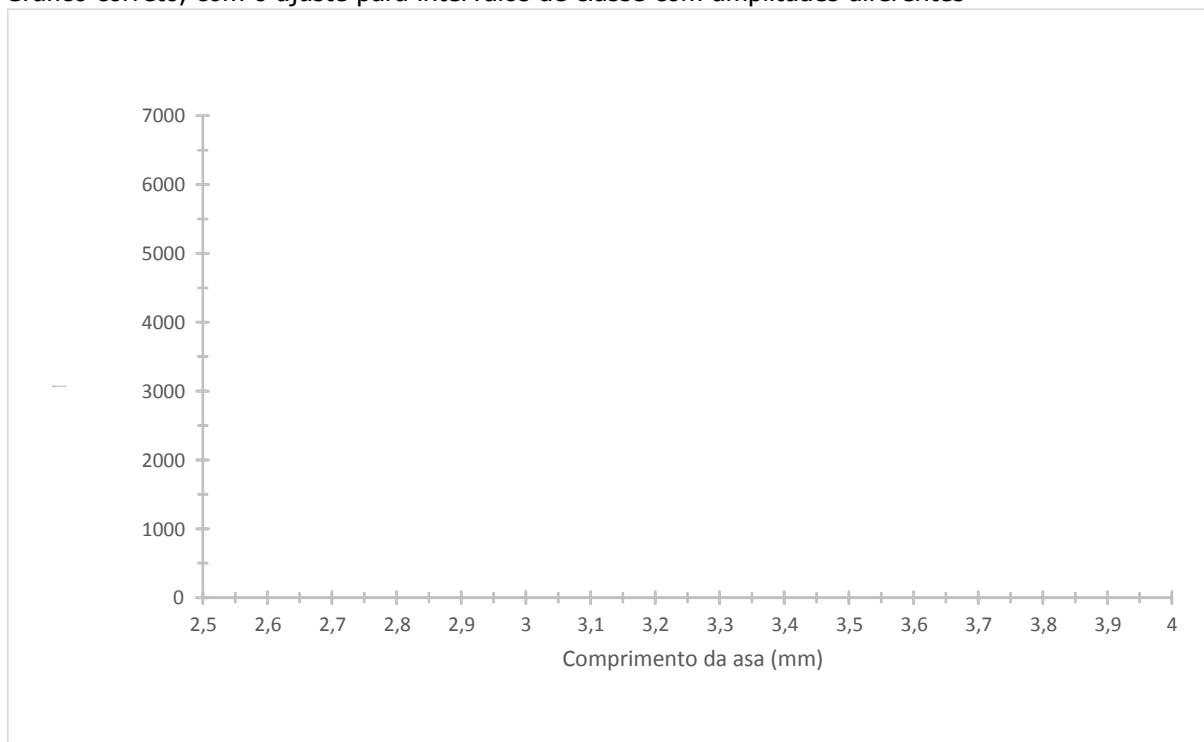


Gráfico correto, com o ajuste para intervalos de classe com amplitudes diferentes



Fonte: Adaptado de Landry SV et al. 1988. Journal of the American Mosquito Control Association, 4(2),1988.

Distribuição de fêmeas de *Aedes triseriatus* segundo comprimento da asa em (mm). Iowa Co.Wisconsin, 1988

Exercício 17

Distribuição de *Culex quinquefasciatus* segundo peso seco (mg). Rio Pinheiros. São Paulo, 2014

Peso seco (mg)	n	%
0,20 -- 0,30	3	15
0,30 -- 0,40	4	20
0,40 -- 0,50	4	20
0,50 -- 0,60	6	30
0,60 -- 0,70	2	10
0,70 -- 0,80	1	5
Total	20	100

Fonte: Marchi MJ. Padronização de técnica para produção em massa de *Culex quinquefasciatus* (Diptera:Culicidae). São Paulo, 2014. [Dissertação de Mestrado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Apresente os dados em um histograma

Exercício 18

Distribuição de espécimes de *Aedes scapularis* segundo o tamanho do corpo. Pariquera-Açu, São Paulo, 2011

Tamanho do centróide (mm)	n	%
1,3 --1,6	11	100
1,6 --1,9	23	100
1,9 --2,2	35	100
2,2 --2,5	32	100
2,5 --2,8	37	100
2,8 --3,1	14	100
Total	152	100

Fonte: Devicari et al. 2013 (dados adaptados)

Apresente os dados em um histograma

Exercício 19

Distribuição de subunidades de coleta de larvas de *Anopheles* spp em áreas da amazônia, segundo valores de ph. Amazônia, 2014

Níveis de Ph	n
4,0 --4,5	9
4,5 --5,0	31
5,0 --5,5	364
5,5 --7,0	92
7,0 --7,5	90
7,5 --7,9	9
Total	595

Fonte: Paulo Rufalco Moutinho. Tese de Doutorado. Perturbações no ambiente natural, e emergência de habitats larvais ocupados por espécies de *Anopheles* Meigen (Diptera: Culicidae) em assentamento rural no sudeste da Amazônia brasileira (adaptado), 2015.

Apresente os dados em um histograma

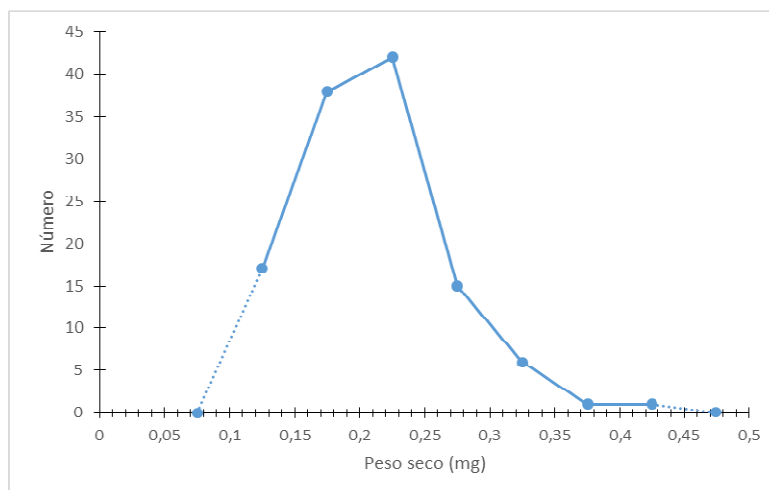
Polígono de frequência simples

Intervalos de classe com mesma amplitude

Distribuição de fêmeas de *Anopheles darlingi*, criadas em laboratório, segundo peso seco (mg). Capanema, Estado do Pará, Brasil. 1995

Classes de peso seco (mg)	n	%
0,10 -- 0,15	17	14,2
0,15 -- 0,20	38	31,7
0,20 -- 0,25	42	35,0
0,25 -- 0,30	15	12,5
0,30 -- 0,35	6	5,0
0,35 -- 0,40	1	0,8
0,40 -- 0,45	1	0,8
Total	120	100

Fonte: Adaptado de Lounibos LP et al., 1995



Fonte: Adaptado de Lounibos LP et al., 1995

Distribuição de fêmeas de *Anopheles darlingi*, criadas em laboratório, segundo peso seco (mg). Capanema, Estado do Pará, Brasil. 1995

Exercício 20

Distribuição de espécimes de *Aedes scapularis* segundo o tamanho do corpo. Pariquera-Açu, São Paulo, 2011

Tamanho do centróide (mm)	n	%
1,3 --1,6	11	100
1,6 --1,9	23	100
1,9 --2,2	35	100
2,2 --2,5	32	100
2,5 --2,8	37	100
2,8 --3,1	14	100
Total	152	100

Fonte: Devicari et al. 2013 (dados adaptados)

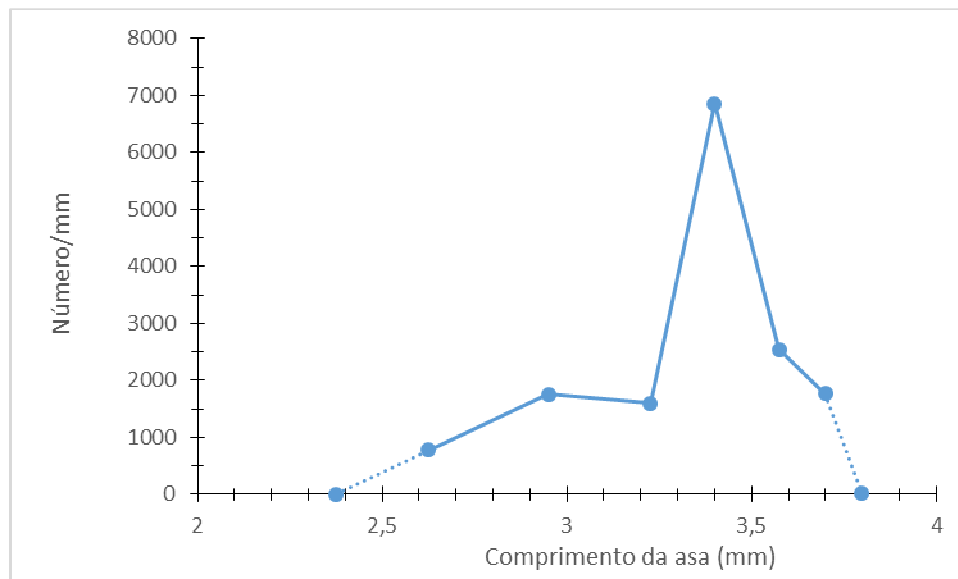
- Apresente a variável em um polígono de frequências simples.
- Interprete os resultados.

Intervalos de classe com amplitudes diferentes

Distribuição de fêmeas de *Aedes triseriatus* segundo comprimento da asa em (mm). Iowa Co. Wisconsin, 1988.

Comprimento da asa (mm)	Nº	Amplitude	Nº/amplitude
2,50 --2,75	195	0,25	780,0
2,75 --3,15	701	0,40	1752,5
3,15 --3,30	240	0,15	1600,0
3,30 --3,50	1371	0,20	6855,0
3,50 --3,65	381	0,15	2540,0
3,65 --3,75	176	0,10	1760,0
Total	3064		

Fonte: Adaptado de Landry SV et al. 1988. Journal of the American Mosquito Control Association, 4(2),1988.



Fonte: Adaptado de Landry SV et al. 1988. Journal of the American Mosquito Control Association, 4(2),1988.
Distribuição de fêmeas de *Aedes triseriatus* segundo comprimento da asa em (mm). Iowa Co. Wisconsin, 1988.

Exercício 21

Distribuição de subunidades de coleta de larvas de *Anopheles* spp em áreas da amazônia, segundo valores de ph. Amazônia, 2014

Ph	n
4,0 --4,5	9
4,5 --5,0	31
5,0 --5,5	364
5,5 --7,0	92
7,0 --7,5	90
7,5 --7,9	9
Total	595

Fonte: Paulo Rufalco Moutinho. Tese de Doutorado. Perturbações no ambiente natural, e emergência de habitats larvais ocupados por espécies de *Anopheles* Meigen (Diptera: Culicidae) em assentamento rural no sudeste da Amazônia brasileira (adaptado), 2015.

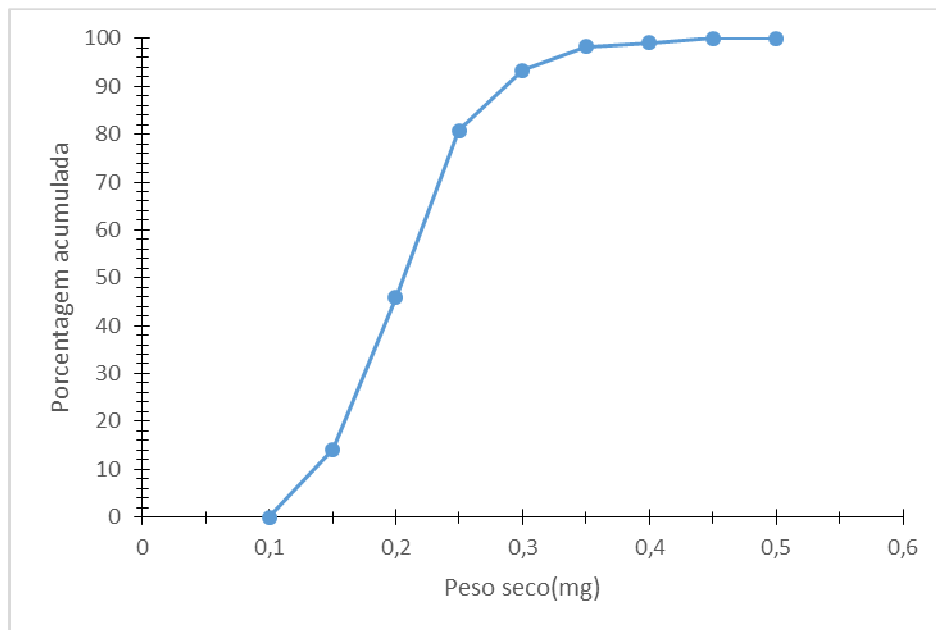
Apresente os dados em um polígono de frequências simples

Polígono (ogiva) de frequências acumuladas

Distribuição de fêmeas de *Anopheles darlingi*, criadas em laboratório, segundo peso seco (mg). Capanema, Estado do Pará, Brasil. 1995

Classes de peso seco (mg)	n	%	% acumulada
0,10 -- 0,15	17	14,2	14,2
0,15 -- 0,20	38	31,7	45,8
0,20 -- 0,25	42	35,0	80,8
0,25 -- 0,30	15	12,5	93,3
0,30 -- 0,35	6	5,0	98,3
0,35 -- 0,40	1	0,8	99,2
0,40 -- 0,45	1	0,8	100,0
Total	120	100	

Fonte: Adaptado de Lounibos LP et al., 1995



Fonte: Adaptado de Lounibos LP et al., 1995

Distribuição acumulada de fêmeas de *Anopheles darlingi*, criadas em laboratório, segundo peso seco (mg). Capanema, Estado do Pará, Brasil. 1995

Percentil	Valor da variável	Medidas estatísticas
25%	0,16 mm	Q1 – primeiro quartil
50%	0,20 mm	Q2 - segundo quartil ou mediana
75%	0,24 mm	Q3 – terceiro quartil

Exercício 22

Distribuição de subunidades de coleta de larvas de *Anopheles* spp em áreas da amazônia, segundo valores de ph. Amazônia, 2014

Ph	n
4,0 --4,5	9
4,5 --5,0	31
5,0 --5,5	364
5,5 --7,0	92
7,0 --7,5	90
7,5 --7,9	9
Total	595

Fonte: Paulo Rufalco Moutinho. Tese de Doutorado. Perturbações no ambiente natural, e emergência de habitats larvais ocupados por espécies de *Anopheles* Meigen (Diptera: Culicidae) em assentamento rural no sudeste da Amazônia brasileira (adaptado), 2015.

- Represente os dados em um polígono de frequências acumuladas.
- Utilizando o gráfico, identifique o valor do ph que deixa 25% das subunidades amostrais abaixo.
- Qual o valor de ph que divide a distribuição em 2 partes iguais, isto é, qual é o valor da variável que deixa 50% das observações abaixo dele?
- Qual a proporção de subunidades amostrais entre 6,0 e 7,0?
- Qual é o valor de ph que deixa 95% das subunidades amostrais abaixo dele?

Representação gráfica de duas variáveis qualitativas

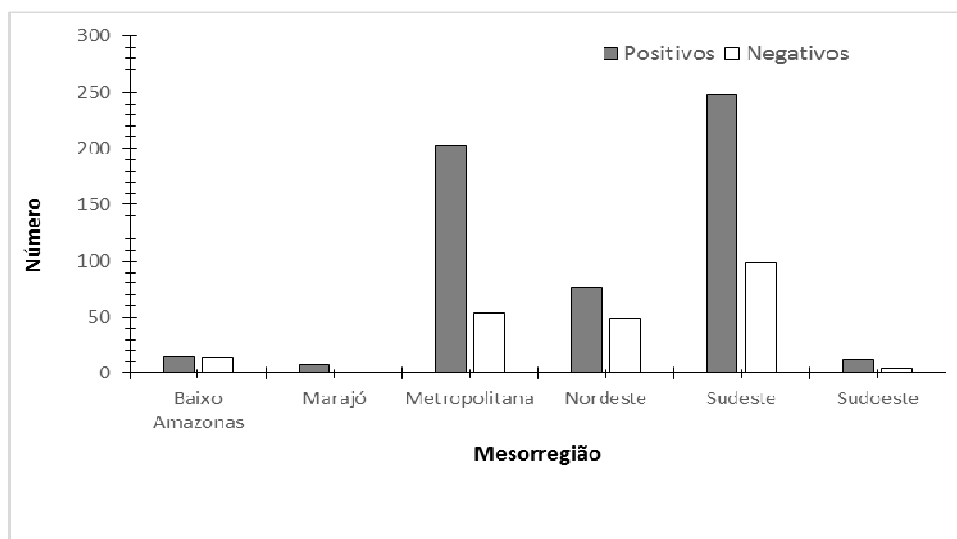
A tabela abaixo foi extraída de estudo que objetivou estudar pacientes com suspeita de dengue ou febre amarela de seis mesorregiões do Estado do Pará.

Tabela 1 - Distribuição da positividade do teste de inibição da hemaglutinação para Flavivirus, por mesorregião, Pará, jun/dez 1999.

Mesorregião	Positivos		Negativos		Total
	n°	%	n°	%	n°
Baixo Amazonas	15	51,7	14	48,3	29
Marajó	8	88,9	1	11,1	9
Metropolitana	203	79,0	54	21,1	257
Nordeste	77	61,1	49	38,9	126
Sudeste	248	71,5	99	28,5	347
Sudoeste	12	70,6	5	29,4	17
Total	563	71,7	222	28,3	785

Fonte: Fichas epidemiológicas – Seção de Arbovírus/IEC, junho a dezembro de 1999

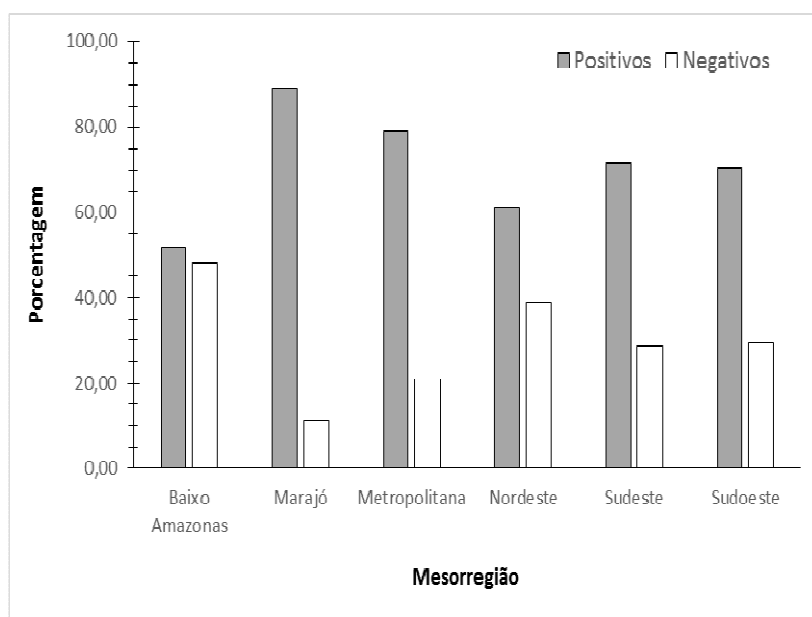
Fonte: Araújo TP et al., 2002. Revista Brasileira de Medicina Tropical 35(6):579-584, 2002.



Fonte: Araújo TP et al., 2002. Revista Brasileira de Medicina Tropical 35(6):579-584, 2002.

Distribuição da positividade de teste de inibição da hemaglutinação para Flavivirus, por mesorregião, Pará, jun/dez 1999

Calculando-se as porcentagens, tomando-se as categorias da variável mesorregião como 100%, tem-se:



Fonte: Araújo TP et al., 2002. Revista Brasileira de Medicina Tropical 35(6):579-584, 2002.

Distribuição da positividade de teste de inibição da hemaglutinação para Flavivirus, por mesorregião, Pará, jun/dez 1999

Exercício 23

Distribuição de fêmeas de *Ae. triseriatus* segundo tamanho do corpo e taxa de paridade, Wisconsin, Estados Unidos, 1985.

Tamanho do corpo (mm)	Paridas		Não paridas		Total	
	n	%	n	%	n	%
Pequeno ($< 2,90$)	23		25		48	
Médio ($2,90 \leq X \leq 3,65$)	127		113		240	
Grande ($\geq 3,66$)	12		28		40	
Total	162		166		328	

Fonte: Landry SV, et.al., 1988. (Adaptado). Journal of the American Mosquito Control Association vol4, nº2.

Apresente os dados em um gráfico considerando como 100% o total nas categorias do tamanho do corpo para investigar a associação entre estas variáveis.

Exercício 24

Considere os dados apresentados na tabela abaixo, coletados em estudo da paridade de fêmeas de *Anopheles cruzii* por período do dia, capturados por aspiração, de dezembro de 2006 a março de 2007 em Paranaguá, Estado do Paraná.

Distribuição do número de *Anopheles cruzii* segundo grau de paridade e período do dia. Paranaguá, Paraná, 2006 e 2007.

Período do dia	Nulípara		Unípara ou Bípara		Total	
	n	%	n	%	n	%
Anterior ao crepúsculo vespertino	27		9		36	
Crepúsculo vespertino	48		14		62	
Posterior ao crepúsculo vespertino	50		17		67	
Total	125		40		165	

Fonte: Adaptado de Bona ACD, Navarro-Silva, MA. Neotropical Entomology 39(2): 282-288 (2010).

Apresente os dados em um gráfico cruzando as duas variáveis para investigar a associação entre elas. Considere como 100% o total nas categorias do período do dia.

Representação gráfica de duas variáveis quantitativas

Exemplo

Os dados a seguir são relativos ao peso seco (mg) de fêmeas de *Culex quinquefasciatus* cujas larvas foram tratadas com mistura de ração de peixe, leite ninho e ração de cão, submetidas a temperaturas de 20°C (*) e acima de 20°C.

0,62*	0,77*	0,84*		0,48	0,61	0,64	0,72
0,64*	0,77*	0,94*		0,50	0,61	0,64	0,72
0,65*	0,79*			0,50	0,62	0,64	0,72
0,70*	0,79*			0,59	0,62	0,66	0,73
0,72*	0,80*			0,59	0,62	0,66	0,73
0,73*	0,80*			0,59	0,62	0,70	0,73
0,73*	0,81*			0,60	0,62	0,70	0,74
0,74*	0,83*			0,61	0,64	0,70	0,75

Fonte: Marchi MJ. Padronização de técnica para produção em massa de *Culex quinquefasciatus* (Diptera: Culicidae). São Paulo, 2014 [Dissertação de Mestrado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

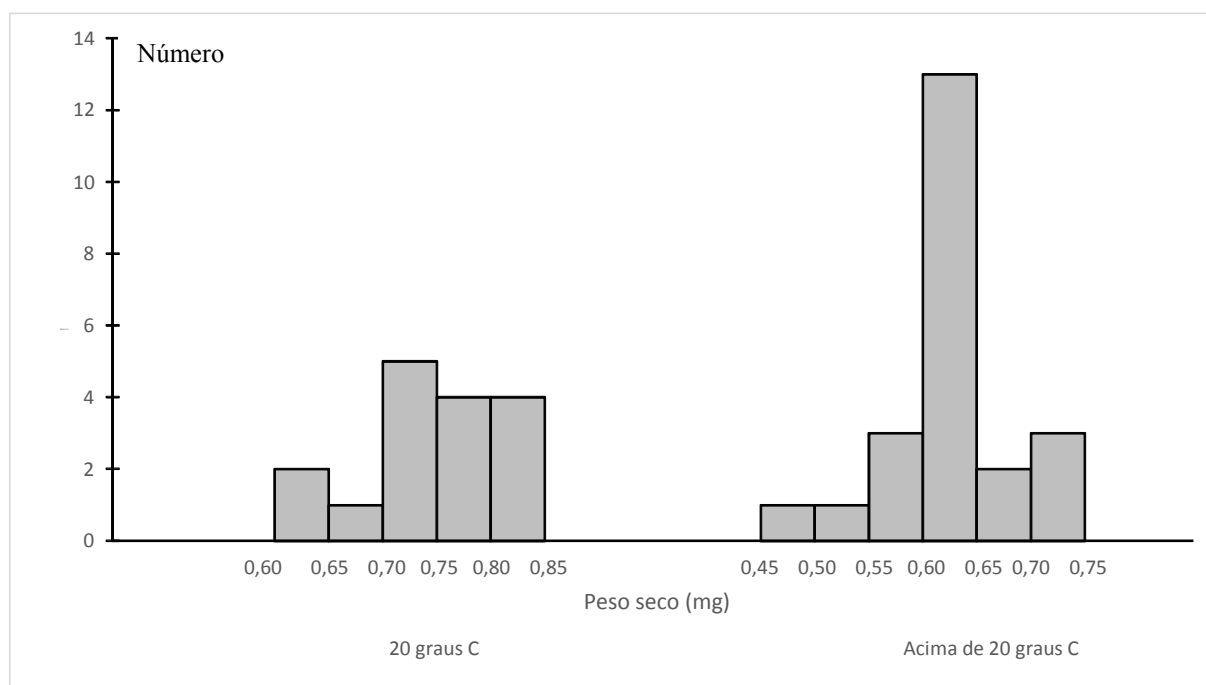
(*) larvas submetidas a temperatura de 20°C.

Distribuição de fêmeas de *Culex quinquefasciatus* segundo peso seco (mg) e temperatura

Peso seco (mg)	20°C		>20°C		Total	
	n	%	n	%	n	%
0,45 -- 0,50	0	-	1	4,35	1	2,56
0,50 -- 0,55	0	-	1	4,35	1	2,56
0,55 -- 0,60	0	-	3	13,04	3	7,69
0,60 -- 0,65	2	12,50	13	56,52	15	38,46
0,65 -- 0,70	1	6,25	2	8,70	3	7,69
0,70 -- 0,75	5	31,25	3	13,04	8	20,51
0,75 -- 0,80	4	25,00	0	-	4	10,26
0,80 -- 0,85	4	25,00	0	-	4	10,26
Total	16	100	23	100	39	100

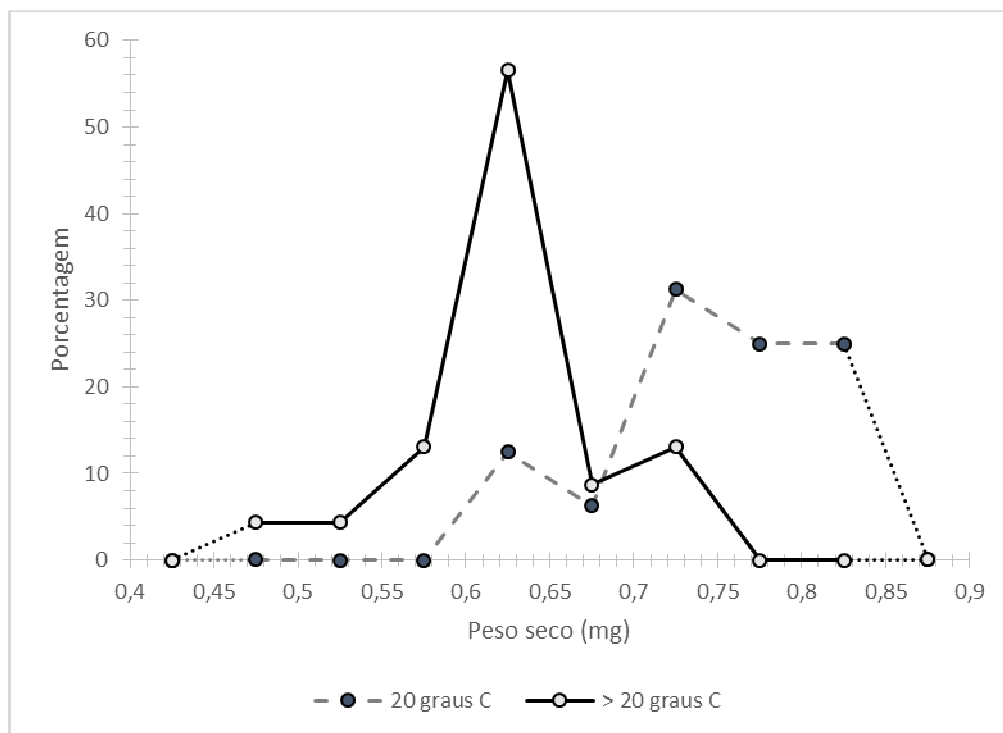
Fonte: Marchi MJ. Padronização de técnica para produção em massa de *Culex quinquefasciatus* (Diptera: Culicidae). São Paulo, 2014 [Dissertação de Mestrado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Distribuição de fêmeas de *Culex quinquefasciatus* segundo peso seco (mg) e temperatura



Fonte: Marchi MJ. Padronização de técnica para produção em massa de *Culex quinquefasciatus* (Diptera: Culicidae). São Paulo, 2014 [Dissertação de Mestrado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Distribuição de fêmeas de *Culex quinquefasciatus* segundo peso seco (mg) e temperatura



Fonte: Marchi MJ. Padronização de técnica para produção em massa de *Culex quinquefasciatus* (Diptera: Culicidae). São Paulo, 2014 [Dissertação de Mestrado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Distribuição de fêmeas de *Culex quinquefasciatus* segundo peso seco (mg) e temperatura

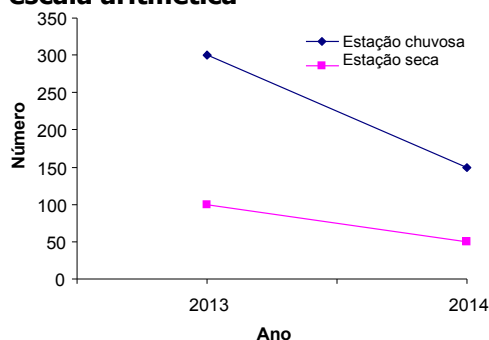
Escala aritmética e escala logarítmica

Número de mosquitos (Culicidae) segundo estação seca e chuvosa. , Habitat X. 2013 e 2014.

Ano	Estação chuvosa	Estação seca
2013	300	100
2014	150	50

Fonte: dados hipotéticos.

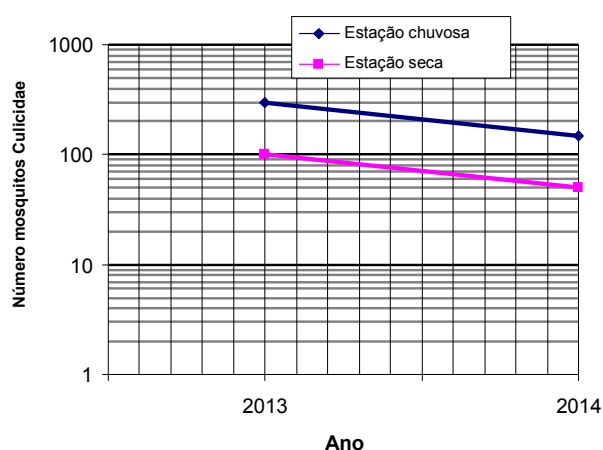
Gráfico em escala aritmética



Fonte: dados hipotéticos

Número de mosquitos Culicidae segundo estação do ano. Habitat X, 2013 e 2014.

Gráfico em escala logarítmica



Fonte: dados hipotéticos.

Número de mosquitos Culicidae segundo estação do ano. Habitat X, 2013 e 2014.

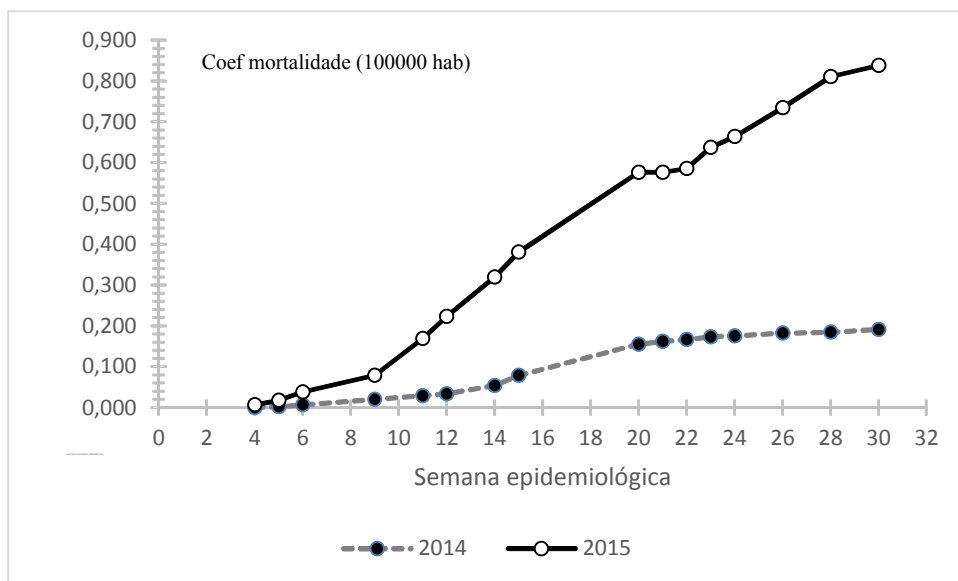
Exemplo

Mortalidade* por dengue segundo semana epidemiológica. Estado de São Paulo**, 2014 e 2015

Semana epidemiológica	Óbitos		Coeficiente*	
	2014	2015	2014	2015
4	0	3	0,000	0,007
5	1	8	0,002	0,018
6	3	17	0,007	0,038
9	9	35	0,020	0,079
11	13	75	0,029	0,169
12	15	99	0,034	0,223
14	24	142	0,054	0,320
15	35	169	0,079	0,381
20	69	256	0,155	0,577
21	72	256	0,162	0,577
22	74	260	0,167	0,586
23	77	283	0,173	0,637
24	78	295	0,176	0,664
26	81	326	0,182	0,734
28	82	360	0,185	0,811
30	85	372	0,191	0,838

* Por 100000 habitantes

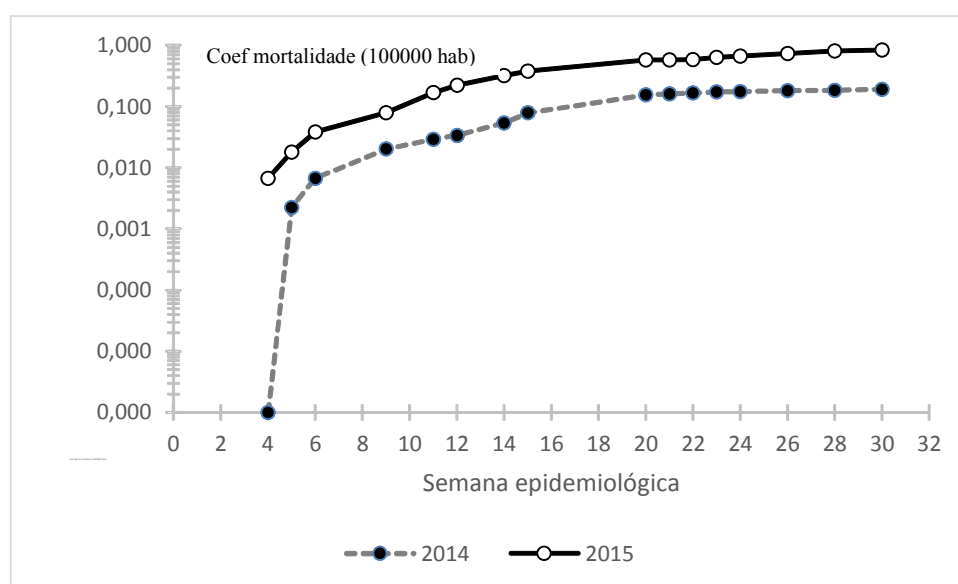
**População do estado de São Paulo estimada (IBGE) para 2015 = 44396484 hab.



Fonte: Boletim Epidemiológico. Secretaria de Vigilância em saúde. MS, 2015

Mortalidade* por dengue segundo semana epidemiológica. Estado de São Paulo, 2014 e 2015

Gráfico em escala logarítmica



Fonte: Boletim Epidemiológico. Secretaria de Vigilância em saúde. MS, 2015

Mortalidade* por dengue segundo semana epidemiológica. Estado de São Paulo, 2014 e 2015

Exemplo

Apresente os dados abaixo graficamente.

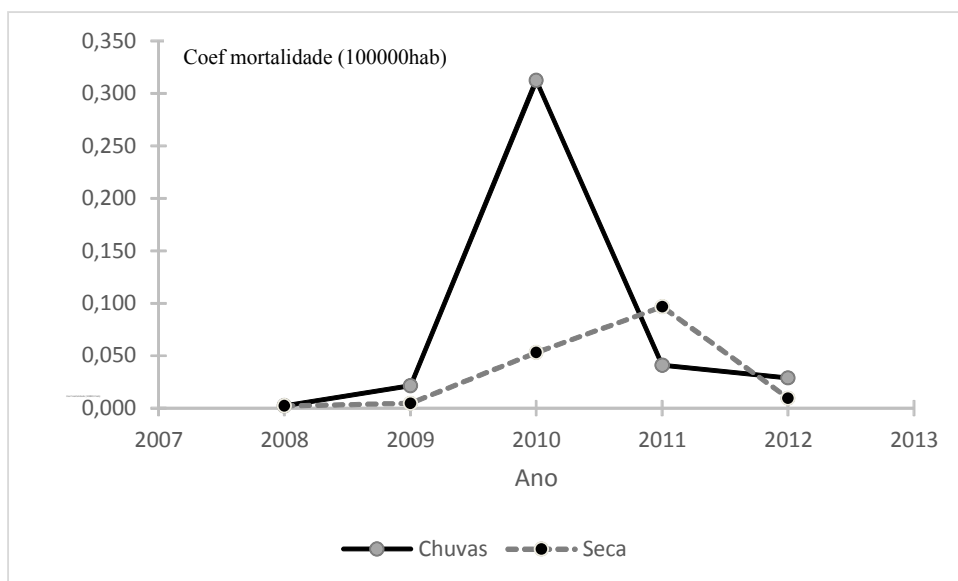
Mortalidade* por dengue segundo período de chuvas e de seca⁽¹⁾. Estado de São Paulo, 2014 e 2015

Ano	Óbitos		Coeficiente	
	Chuvas	Seca	Chuvas	Seca
2008	1	1	0,002	0,002
2009	9	2	0,022	0,005
2010	129	22	0,313	0,053
2011	17	40	0,041	0,097
2012	12	4	0,029	0,010

* Por 100000 habitantes

(1) chuvas: outubro a março; seca: abril a setembro

**População do estado de São Paulo estimada (IBGE) para 2010 = 41262199 hab

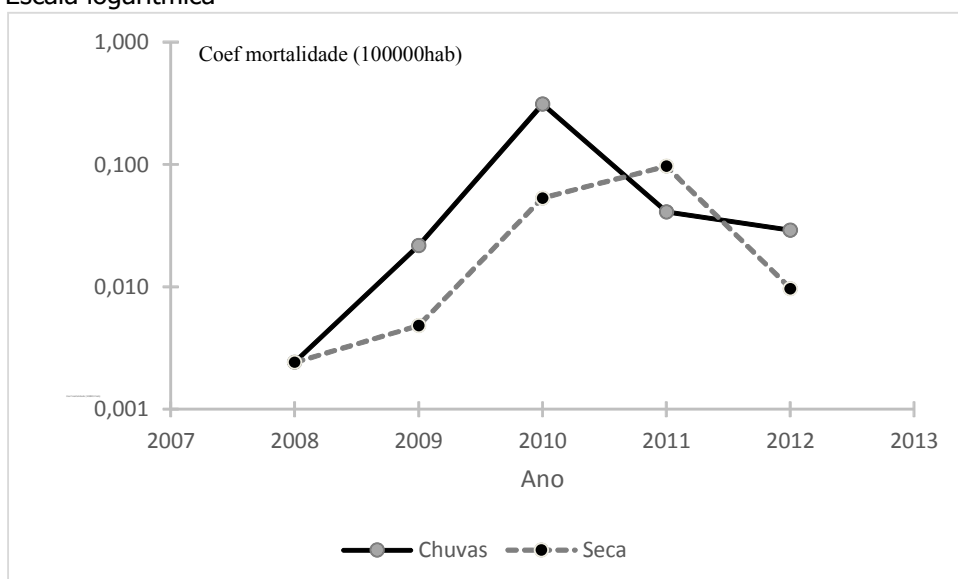


Fonte: Boletim Epidemiológico. Secretaria de Vigilância em saúde. MS, 2015

* Por 100000 habitantes; (1) chuvas: outubro a março; seca: abril a setembro

Mortalidade* por dengue segundo período de chuvas e de seca. Estado de São Paulo, 2014 e 2015

Escala logaritmica



Exercício 25

Os dados são referentes a taxa de incidência de dengue (número de casos confirmados, por 100 mil habitantes) nos Estados de São Paulo e Rio de Janeiro, Região Sudeste do Brasil. Período de 2004 a 2014

Ano	São Paulo	Rio de Janeiro
2004	7,8	8,2
2005	14,3	8,9
2006	130,8	171,0
2007	221,6	367,7
2008	17,9	1249,1
2009	21,8	42,5
2010	503,0	186,5
2011	278,4	1036,8
2012	69,6	1116,2
2013	506,0	1301,6
2014	515,2	46,9

Fonte: SES/SINAN

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

- Apresente os dados em um gráfico utilizando escala logarítmica.
- Interprete os resultados.

Exercício 26

Distribuição mensal do número de *Culex quinquefasciatus* segundo sexo. São Paulo, novembro de 2003 a março de 2004.

Mês/Ano	Macho	Fêmea
Nov/2003	644	168
Dez/2003	443	112
Jan/2004	1198	284
Fev/2004	192	87
Mar/2004	53	21

Fonte: Adaptado de Laporta et.al.2006. Revista Brasileira de Entomologia 50(1):125-127, março 2006.

- Apresente os dados em um gráfico utilizando escala logarítmica.
- Interprete os resultados.

Medidas de tendência central (média e mediana)

Medidas de tendência central

Média aritmética

Notação:

$X \rightarrow$ variável

$N \rightarrow$ tamanho da população

$n \rightarrow$ tamanho da amostra

$\mu \rightarrow$ Média populacional (parâmetro, geralmente desconhecido)

$\bar{X} \rightarrow$ Estatística (fórmula)

$\bar{x} \rightarrow$ Média amostral (estimativa, valor calculado na amostra)

Média aritmética

Considerar

X: Número de ovos de *Aedes aegypti*

3	2	5	6	4
---	---	---	---	---

Para calcular a média soma-se os valores de uma variável e divide-se a soma pelo número de valores.

$$\text{Média aritmética} = \frac{3+2+5+6+4}{5} = 4 \text{ ovos}$$

Ordenando-se os valores,

2	3	4	5	6
média				

2-4=	-2
3-4=	-1
4-4=	0
5-4=	1
6-4=	2
Soma=	0

Média aritmética é o valor que indica o centro de equilíbrio de uma distribuição de frequências de uma variável quantitativa. Portanto, a soma das diferenças entre cada valor e a média é igual a zero.

Apresentação em fórmula

Em uma amostra aleatória simples de tamanho n , composta pelas observações x_1, x_2, \dots, x_n , a média aritmética (\bar{x}) é igual a:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}$$

No exemplo, $x_1=3; x_2=2, x_3=5, x_4=6, x_5=4; n=5$. Portanto, $\bar{x} = \frac{3+2+5+6+4}{5} = \frac{20}{5} = 4 \text{ ovos}$

OBS: a média aritmética

- só existe para variáveis quantitativas e seu valor é único;
- é da mesma natureza da variável considerada;
- sofre influência dos valores aberrantes (outlier)

Ex: $x_1=3; x_2=2, x_3=5, x_4=6, x_5=24; n=5$. Portanto, $\bar{x} = \frac{3+2+5+6+24}{5} = \frac{40}{5} = 8 \text{ ovos}$

Exemplo:

Os dados a seguir são relativos à quantidade mensal de larvas de *Aedes albopictus* coletadas em dois ambientes do Parque Ecológico do Tietê, Guarulhos, SP, no período de abril de 2001 a março de 2003. Os dados foram extraídos de Urbinatti PR. "Observações ecológicas de *Aedes albopictus* (Diptera: Culicidae) em áreas de proteção ambiental e urbana da periferia na Grande São Paulo. São Paulo, 2004". [Tese de Doutorado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Ambiente A

111	117	170	113	163	212	173	155	114	167
109	158	220	118	129	112	130	128	135	119

Número médio de larvas de *Ae. albopictus*:

$$\bar{x}_A = \frac{111 + 117 + \dots + 135 + 119}{20} = 142,7 \text{ larvas}$$

Ambiente B

118	105	159	113	149	72	76	83	92	104
137	84	87	158	138	137	130	112	122	142

$$\bar{x}_B =$$

Exercício 27

Os dados a seguir são provenientes de um estudo que avaliou o tempo médio de vida em dias de 22 machos e 31 fêmeas de *Triatoma sordida*, nos estágios de ninfa e adulto, em condições de laboratório.

Utilizou-se neste exemplo apenas os dados de tempo de vida em estágio de ninfa.

Calcule o número médio de dias no estágio de ninfa para machos:

Machos

136	157	154	129	247	164	133	126	247	139
139	148	221	248	131	139	135	143	249	173
241	241								

$$\bar{x}_{Machos} =$$

Calcule o número médio de dias no estágio de ninfa para fêmeas:

Fêmeas

126	126	127	130	129	128	131	126	132	136
146	128	150	136	158	134	126	128	128	139
203	208	242	241	250	244	259	241	253	234
250									

$$\bar{x}_{Fêmeas} =$$

Fonte: Souza JMP de, 1978. *Triatoma sordida* – Considerações sobre o tempo de vida das formas adultas e sobre a oviposição das fêmeas. Revista de Saúde Pública. São Paulo, 12:291-6.

Média geométrica

É a raiz n-ésima do produto de n observações $\bar{X}_G = \sqrt[n]{X_1 X_2 X_3 \dots X_n} = \sqrt[n]{\prod_{i=1}^n X_i}$

A média geométrica também pode ser calculada como o anti logaritmo da média aritmética dos logaritmos dos valores, onde o logaritmo pode estar em qualquer base.

$$\bar{X}_G = \text{anti log}\left(\frac{\log X_1 + \log X_2 + \dots + \log X_n}{n}\right) = \frac{\sum_{i=1}^n \log X_i}{n}$$

É apropriada somente para valores positivos. Se os valores forem todos iguais, a média aritmética e a geométrica serão idênticas, caso contrário, $\bar{X}_G < \bar{X}$. É útil para razões onde se deseja dar pesos iguais a cada razão e calculando a média para mudanças percentuais.

Exemplo:

X: Número de ovos de *Aedes aegypti*

3	2	5	6	4
---	---	---	---	---

$$\bar{x}_G = \sqrt[5]{3 \times 2 \times 5 \times 6 \times 4} = \sqrt[5]{\prod_{i=1}^5 X_i} = \sqrt[5]{720} = 3,7 \text{ ovos}$$

ou

$$\bar{x}_G = \text{anti log}\left(\frac{\log 3 + \log 2 + \dots + \log 6}{5}\right) = \frac{\sum_{i=1}^5 \log X_i}{5}$$

Valor	Log(valor)	Ln(valor)
3	0,477	1,099
2	0,301	0,693
5	0,699	1,609
6	0,778	1,792
4	0,602	1,386
Soma	2,857	6,579

$$\bar{x}_G = \text{anti log}\left(\frac{2,857}{5}\right) = \text{anti log}(0,5714) = 3,7 \text{ ovos}$$

$$\bar{x}_G = \text{anti ln}\left(\frac{6,579}{5}\right) = \text{anti ln}(1,3158) = 3,7 \text{ ovos}$$

$$\bar{x} = 4,0 \text{ ovos}$$

Exercício 28

Considere as observações, calcule e compare as medidas de resumo

3	2	5	6	47
---	---	---	---	----

Mediana

É o valor que ocupa a posição central de uma série de n observações, quando estas estão ordenadas de forma crescente ou decrescente.

Quando número de observações (n) for **ímpar**:

a mediana é o valor da variável que ocupa o posto $\frac{n+1}{2}$

Quando o número de observações (n) for **par**:

a mediana é a média aritmética dos valores da variável que ocupam os postos $\frac{n}{2}$ e $\frac{n+2}{2}$

OBS:

- existe para variável quantitativa e qualitativa ordinal;
- é da mesma natureza da variável considerada;
- torna-se inadequada quando há muitos valores repetidos;
- não sofre influência de valores aberrantes.

Exemplo:

Os dados a seguir são relativos à quantidade mensal de larvas de *Aedes albopictus* coletadas em dois ambientes do Parque Ecológico do Tietê, Guarulhos, SP, no período de abril de 2001 a março de 2003. Os dados foram extraídos de Urbinatti PR. "Observações ecológicas de *Aedes albopictus* (Diptera:Culicidae) em áreas de proteção ambiental e urbana da periferia na Grande São Paulo. São Paulo, 2004". [Tese de Doutorado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Ambiente A

111	117	170	113	163	212	173	155	114	167
109	158	220	118	129	112	130	128	135	119

Ordenando-se os valores:

109	112	114	118	128	130	155	163	170	212
111	113	117	119	129	135	158	167	173	220

Número mediano de larvas de *Ae.albopictus*:

Valor mediano: $(129+130)/2 = 129,5$ larvas

Ambiente B

118	105	159	113	149	72	76	83	92	104
137	84	87	158	138	137	130	112	122	142

Ordenando-se os valores:

Número mediano de larvas de *Ae.albopictus*:

Exercício 29

Os dados a seguir são provenientes de um estudo que avaliou o tempo médio de vida em dias de 22 machos e 31 fêmeas de *Triatoma sordida*, nos estágios de ninfa e adulto, em condições de laboratório (Souza JMP de, 1978. *Triatoma sordida* – Considerações sobre o tempo de vida das formas adultas e sobre a oviposição das fêmeas. Revista de Saúde Pública. São Paulo, 12:291-6).

Utilizou-se neste exemplo apenas os dados de tempo de vida em estágio de ninfa.

Calcule o número mediano de dias no estágio de ninfa para machos:

Machos

136	157	154	129	247	164	133	126	247	139
139	148	221	248	131	139	135	143	249	173
241	241								

Ordenando-se os valores

Valor mediano=

Número mediano de dias no estágio de ninfa para fêmeas:

Fêmeas

126	126	127	130	129	128	131	126	132	136
146	128	150	136	158	134	126	128	128	139
203	208	242	241	250	244	259	241	253	234
250									

valor mediano=

Ordenando-se os valores

Valor mediano=

Medidas de dispersão (variância, desvio-padrão, coeficiente de variação e percentis)

Constituem medidas de dispersão

- Valores mínimo e máximo
- Amplitude de variação
- Variância
- Desvio padrão
- Coeficiente de variação de Pearson

Valores mínimo e máximo: valores extremos da distribuição.

Ambiente A

109	112	114	118	128	130	155	163	170	212
111	113	117	119	129	135	158	167	173	220

Valor mínimo = 109 larvas; valor máximo = 220 larvas

Ambiente B

118	105	159	113	149	72	76	83	92	104
137	84	87	158	138	137	130	112	122	142

Ordenando-se os valores:

72	76	83	84	87	92	104	105	112	113
118	122	130	137	137	138	142	149	158	159

Valor mínimo = 72 larvas; valor máximo = 159 larvas

Amplitude de variação: é a diferença entre os 2 valores extremos da distribuição.

Ambiente A

Valor máximo - valor mínimo = $220 - 109 = 111$ larvas

Ambiente B

Valor máximo - mínimo = $159 - 72 = 87$ larvas

Variância

É uma medida de dispersão que fornece a distância média ao quadrado das observações em relação à média. As distâncias de cada observação em relação à média são denominadas desvios em relação à média. Se forem elevados ao quadrado, são denominados desvios quadráticos. Então a variância também pode ser entendida como a média dos desvios quadráticos de cada observação em relação à média aritmética.

Considerar os valores

3	2	5	6	4
---	---	---	---	---

$$\bar{x} = 4 \text{ ovos}$$

Valor	(valor-média)	(valor-média)	(valor-média) ²
3	3-4=	-1 ovos	1 ovos ²
2	2-4=	-2 ovos	4 ovos ²
5	5-4=	1 ovos	1 ovos ²
6	6-4=	2 ovos	4 ovos ²
4	4-4=	0 ovos	0 ovos ²
Soma =		0 ovos	10 ovos ²

2	3	4	5	6
---	---	---	---	---

$$\text{Variância} = \frac{10}{5} = 2 \text{ ovos}^2$$

Desvio padrão

É uma medida de dispersão calculada a partir da variância sendo a raiz quadrada desta. Indica o quanto "erramos em média" ao representarmos um conjunto de dados pela média. É portanto, o desvio médio dos valores em relação à média

$$\text{Desvio padrão} = \sqrt{2} = 1,4 \text{ ovos}$$

O erro médio que se comete ao resumir os dados pela média é de 1,4 ovos.

Apresentando as fórmulas:

Na população a variância é representada pelo parâmetro σ^2 que pode ser estimado por dois estimadores:

$$\text{Se os dados forem referentes à toda a população, o estimador é } S_{(N)}^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$$

É a soma dos desvios quadráticos dos valores em relação à média dividida por N, onde N é o número de observações

$$\text{Se os dados forem referentes a uma amostra, o estimador é } S_{(N-1)}^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$$

É a soma dos desvios quadráticos dos valores em relação à média dividida por N-1, onde N é o número de observações

Desvio padrão

Na população, o desvio padrão é um parâmetro com notação σ sendo igual à a raiz quadrada da variância, ou seja $\sigma = \sqrt{\sigma^2}$.

O estimador do desvio padrão é representado por $S = \sqrt{S^2}$

Notação, resumo:

Estatística	População Parâmetro	Estimador	Estimativa (com dados da amostra)
Média	μ	$\bar{X} = \frac{\sum X_i}{N}$	$\bar{x} = \frac{\sum x_i}{n}$
Variância	σ^2	$S_{(N)}^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$	$s_{(N)}^2 = \frac{\sum_{i=1}^N (x_i - \bar{X})^2}{N}$
		$S_{(N-1)}^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$	$s_{(n-1)}^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n-1}$
Desvio padrão	σ	$S = \sqrt{S^2}$	$s = \sqrt{s^2}$

Exemplo

X: número de larvas de *Ae.albopictus*

Ambiente A:

111	117	170	113	163	212	173	155	114	167
109	158	220	118	129	112	130	128	135	119

$$\text{Variância: } s^2 = \frac{(111-142,7)^2 + \dots + (111-142,7)^2}{19} = 1109,19 \text{ larvas}^2$$

$$\text{Desvio padrão } s = \sqrt{1109,19} = 33,30 \text{ larvas}$$

Exercício 30

X: número de larvas de *Ae.albopictus*

Ambiente B:

118	105	159	113	149	72	76	83	92	104
137	84	87	158	138	136	130	112	122	142

Variância:

Desvio padrão

Coefficiente de Variação de Pearson (CV):

É uma medida de dispersão que relaciona a média e o desvio padrão. É representado em porcentagem. Será próximo de zero quando a dispersão for pequena, próxima a zero. Pode ser maior do que 100%. Isto ocorrerá quando a dispersão for maior que a média. O CV não é definido para $\bar{x} = 0$

$$CV = \frac{S}{\bar{X}} \times 100, \text{ onde } S \text{ é o desvio padrão e } \bar{X}, \text{ a média.}$$

Exemplo
Ambiente A

$$\text{Coeficiente de Variação de Pearson } CV = \frac{33,30}{142,6} \times 100 = 23,3\%$$

Exercício 31

Calcule e interprete o coeficiente de variação do número de larvas de *Aedes albopictus* para o ambiente B

Exercício 32

Os dados a seguir são provenientes de um estudo que avaliou o tempo médio de vida (dias) de 22 machos e 31 fêmeas de *Triatoma sordida*, nos estágios de ninfa e adulto, em condições de laboratório (Souza JMP de, 1978. *Triatoma sordida* – Considerações sobre o tempo de vida das formas adultas e sobre a oviposição das fêmeas. Revista de Saúde Pública. São Paulo, 12:291-6).

Utilizou-se neste exemplo apenas os dados de tempo de vida em estágio de ninfa.

Calcule a variância, o desvio padrão e o coeficiente de variação de Pearson para os machos e fêmeas. Compare e discuta os resultados.

Machos

136	157	154	129	247	164	133	126	247	139
139	148	221	248	131	139	135	143	249	173
241	241								

Fêmeas

126	126	127	130	129	128	131	126	132	136
146	128	150	136	158	134	126	128	128	139
203	208	242	241	250	244	259	241	253	234
250									

Quartil

Valores da variável que dividem a distribuição em quatro partes iguais.

$\frac{1}{4}$		$\frac{1}{2}$		$\frac{3}{4}$	
25%		25%		25%	25%

Q1: deixa abaixo 25% das observações

25%	75%
-----	-----

Q2: deixa abaixo 50% das observações

50%	50%
-----	-----

Q3: deixa abaixo 75% das observações

75%	25%
-----	-----

$$Q_1 = x_{\left(\frac{1}{4}(n+1)\right)} \quad \text{e} \quad Q_3 = x_{\left(\frac{3}{4}(n+1)\right)}$$

onde x é o valor da variável e $\left(\frac{1}{4}(n+1)\right)$ e $\left(\frac{3}{4}(n+1)\right)$ são índices que representam as posições ocupadas por x.

Exemplo

Os dados abaixo são referentes ao tamanho da asa (mm), de 50 fêmeas de *Ae. scapularis* segundo estado de paridade. 23 fêmeas eram nulíparas e 27 eram paridas (*).

3,719*	2,991*	3,168*	2,597	3,720
3,365*	2,715*	3,562*	3,404	3,670
3,581*	2,942*	3,453*	2,578	3,719
3,070*	2,627*	3,335*	3,434	3,079
3,603*	2,883*	3,129*	2,784	3,178
3,257*	2,696*	3,316*	2,735	3,168
3,072*	2,873*	2,885*	3,365	3,788
2,628*	2,676*	3,207	2,725	2,450
3,169*	3,217*	3,847	2,930	3,099
2,932*	3,325*	3,503	3,138	2,932

Ordenando-se os dados, em cada grupo, obtém-se:

2,627*	2,991*	3,335*	2,725	3,207
2,628*	3,070*	3,365*	2,735	3,365
2,676*	3,072*	3,453*	2,784	3,404
2,696*	3,129*	3,562*	2,930	3,434
2,715*	3,168*	3,581*	2,932	3,503
2,873*	3,169*	3,603*	3,079	3,670
2,883*	3,217*	3,719*	3,099	3,719
2,885*	3,257*	2,450	3,138	3,720
2,932*	3,316*	2,578	3,168	3,788
2,942*	3,325*	2,597	3,178	3,847

Fonte: Menezes, RMT. Aspectos Ecológicos de *Culex quinquefasciatus* e *Ocherotatus scapularis* abrigados em habitats do Parque Ecológico do Tietê na cidade de São Paulo. São Paulo, 2002. [Tese de Doutorado apresentada a Faculdade de Saúde Pública da Universidade de São Paulo].

Entre as fêmeas de *Aedes scapularis* que são nulíparas:

$$Q_1 = x_{\left(\frac{1}{4}(23+1)\right)} = x_6 = 2,784mm ; \quad Q_3 = x_{\left(\frac{3}{4}(23+1)\right)} = x_{18} = 3,503mm$$

$$Q_2 = x_{\left(\frac{1}{2}(23+1)\right)} = x_{12} = 3,168mm$$

Entre as fêmeas de *Ae. scapularis* que são paridas

$$Q_1 = x_{\left(\frac{1}{4}(27+1)\right)} = x_7 = 2,883mm ; \quad Q_3 = x_{\left(\frac{3}{4}(27+1)\right)} = x_{21} = 3,335mm$$

$$Q_2 = x_{\left(\frac{1}{2}(27+1)\right)} = x_{14} = 3,129mm$$

Se o resultado for um valor fracionário:

Por exemplo, para n=22

Entre nulíparas

$$Q_1 = x_{\left(\frac{1}{4}(22+1)\right)} = x_{\left(\frac{23}{4}\right)} = x_{\left(5\frac{3}{4}\right)}$$

que é $\frac{3}{4}$ do caminho entre $x_5=2,735$ e $x_6=2,784$

$$Q_1 = 2,735 + \frac{3}{4}(2,784 - 2,735) = 2,772mm$$

$$Q_3 = x_{\left(\frac{3}{4}(22+1)\right)} = x_{\left(17\frac{1}{4}\right)}$$

que é $\frac{1}{4}$ do caminho entre $x_{17}= 3,434$ e $x_{18}=3,503$

$$Q_3 = 3,434 + \frac{1}{4}(3,503 - 3,434) = 3,451mm$$

Decil

Valores da variável que dividem a distribuição em dez partes iguais.

Percentil

Valores da variável que dividem a distribuição em cem partes iguais.

Entre as fêmeas de *Aedes scapularis* que são nulíparas

2,627*	2,991*	3,335*	2,725	3,207
2,628*	3,070*	3,365*	2,735	3,365
2,676*	3,072*	3,453*	2,784	3,404
2,696*	3,129*	3,562*	2,930	3,434
2,715*	3,168*	3,581*	2,932	3,503
2,873*	3,169*	3,603*	3,079	3,670
2,883*	3,217*	3,719*	3,099	3,719
2,885*	3,257*	2,450	3,138	3,720
2,932*	3,316*	2,578	3,168	3,788
2,942*	3,325*	2,597	3,178	3,847

(*) não nula para (paridas)

Percentil 5:

$$P_5 = x_{\left(\frac{5}{100}(23+1)\right)} = x_{\left(\frac{120}{100}\right)} = x_{\left(1\frac{1}{5}\right)}$$

que é 1/5 do caminho entre $x_1=2,450$ e $x_2=2,578$

$$P_5 = 2,450 + \frac{1}{5}(2,578 - 2,450) = 2,476mm$$

Percentil 10:

$$P_{10} = x_{\left(\frac{10}{100}(23+1)\right)} = x_{\left(\frac{240}{100}\right)} = x_{\left(2\frac{2}{5}\right)}; P_{10} = 2,578 + \frac{2}{5}(2,597 - 2,578) = 2,586mm$$

Percentil 50:

$$P_{50} = x_{\left(\frac{50}{100}(23+1)\right)} = x_{\left(\frac{1200}{100}\right)} = x_{(12)}; P_{50} = 3,168mm$$

Percentil 75:

$$P_{75} = x_{\left(\frac{75}{100}(23+1)\right)} = x_{\left(\frac{1800}{100}\right)} = x_{(18)}; P_{75} = 3,503mm$$

Percentil 90:

$$P_{90} = x_{\left(\frac{90}{100}(23+1)\right)} = x_{\left(\frac{2160}{100}\right)} = x_{\left(21\frac{3}{5}\right)}; P_{90} = 3,720 + \frac{3}{5}(3,788 - 3,720) = 3,761mm$$

Percentil 95:

$$P_{95} = x_{\left(\frac{95}{100}(23+1)\right)} = x_{\left(\frac{2280}{100}\right)} = x_{\left(22\frac{4}{5}\right)}; P_{95} = 3,788 + \frac{4}{5}(3,847 - 3,788) = 3,835mm$$

Box plot

O Box plot representa graficamente dados de forma resumida em um retângulo onde as linhas da base e do topo são o primeiro e o terceiro quartis, Q1 e Q3 respectivamente. A linha entre Q1 e Q3 é a mediana (Q2). Linhas verticais que iniciam no meio da base e do topo do retângulo, terminam em valores denominados adjacentes inferior e superior (Chambers *et al.*, 1983, pag 60).

O valor adjacente superior (VAS) é o maior valor das observações que é menor ou igual a $Q3 + 1,5(Q3 - Q1)$.

O valor adjacente inferior (VAI) é definido como o menor valor das observações que é maior ou igual a $Q1 - 1,5(Q3 - Q1)$.

A diferença $(Q3 - Q1)$ é denominada intervalo inter-quartil (IIQ).

Após calcular os quartis Q1, Q2 e Q3 e os valores VAS e VAI, é possível desenhar o box plot e se entre as observações restar alguma que caia fora dos valores adjacentes, então estas são chamadas valores outliers.

Assim, *outliers* (discrepantes ou aberrantes) são valores que “fogem” da distribuição dos dados. O box plot além dos outliers permitir investigar a dispersão dos dados. Isto é feito de modo exploratório comparando-se as distâncias entre os quartis Q3 e Q1 e as distâncias entre os VAS e VAI.

Exemplo:

Ambiente A: número de larvas de *Ae.albopictus*

109	112	114	118	128	130	155	163	170	212
111	113	117	119	129	135	158	167	173	250

Ambiente B: número de larvas de *Ae.albopictus*

12	83	87	104	112	118	130	137	142	158
76	84	92	105	113	122	136	138	149	159

Ambiente A:

$n=20$;

$$Q1 = x_{\frac{1}{4}(n+1)} = x_{\frac{21}{4}} = x_{5\frac{1}{4}} = 114 + \frac{1}{4}(117 - 114) = 114 + 0,75 = 114,75$$

$$Q2 = x_{\frac{2}{4}(n+1)} = x_{\frac{42}{4}} = x_{10\frac{1}{2}} = 129 + \frac{1}{2}(130 - 129) = 129 + 0,5 = 129,5$$

$$Q3 = x_{\frac{3}{4}(n+1)} = x_{\frac{63}{4}} = x_{15\frac{3}{4}} = 163 + \frac{3}{4}(167 - 163) = 163 + 3,0 = 166$$

Intervalo Inter-Quartil (IIQ): $Q3 - Q1 = 166 - 114,75 = 51,25$

Valor adjacente superior (VAS)

Primeiro faz-se a conta: $166 + 1,5 \times 51,25 = 242,875$

O valor adjacente superior é o maior valor da distribuição que é menor ou igual à 242,9, portanto VAS = 212

Valor adjacente inferior (VAI)

Primeiro faz-se a conta: $114,75 - 1,5 \times 51,25 = 37,875$

O valor adjacente inferior é o menor valor da distribuição que é maior ou igual à 37,9.

Portanto VAI = 109.

Ambiente B:

n=20

$$Q1 = x_{\frac{1}{4}(n+1)} = x_{\frac{21}{4}} = x_{5\frac{1}{4}} = 87 + \frac{1}{4}(92 - 87) = 87 + 1,25 = 88,25$$

$$Q2 = x_{\frac{2}{4}(n+1)} = x_{\frac{42}{4}} = x_{10\frac{1}{2}} = 113 + \frac{1}{2}(118 - 113) = 113 + 0,5 = 115,5$$

$$Q3 = x_{\frac{3}{4}(n+1)} = x_{\frac{63}{4}} = x_{15\frac{3}{4}} = 137 + \frac{3}{4}(138 - 137) = 137 + 0,75 = 137,75$$

Intervalo Inter-Quartil (IIQ): $Q3 - Q1 = 49,5$

Valor adjacente superior (VAS)

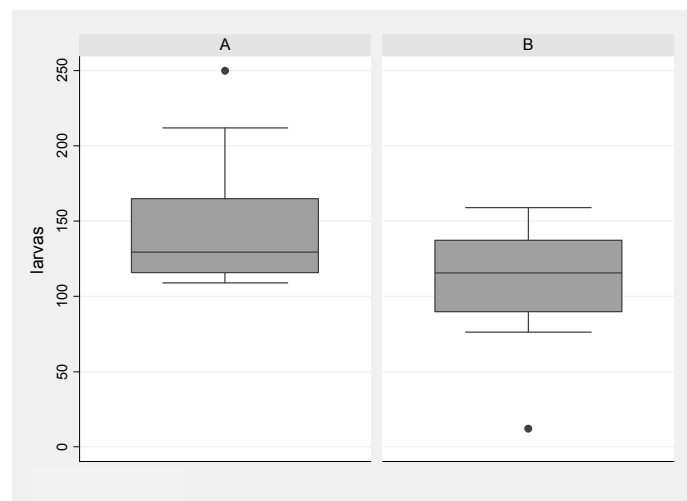
Primeiro faz-se a conta: $137,75 + 1,5 \times 49,5 = 212$

O valor adjacente superior é o maior valor da distribuição que é menor ou igual à 212, portanto VAS = 159

Valor adjacente inferior (VAI)

Primeiro faz-se a conta: $88,25 - 1,5 \times 49,5 = 14$

O valor adjacente inferior é o menor valor da distribuição que é maior ou igual à 14. Portanto VAI = 76.



Fonte: Urbinatti PR. "Observações ecológicas de *Aedes albopictus* (Diptera:Culicidae) em áreas de proteção ambiental e urbana da periferia na Grande São Paulo. São Paulo, 2004". [Tese de Doutorado, Faculdade de Saúde Pública da Universidade de São Paulo]. (Adaptado).

Gráfico - Box plot da variável número de larvas de *A. albopictus*. Áreas de proteção ambiental e urbana da periferia na Grande São Paulo. São Paulo, 2004

Exercício 33

Os dados a seguir são adaptados de artigo publicado por Honório NA & Lourenço-de-Oliveira R. 2001, cujo estudo avaliou a frequência mensal de larvas e pupas de *Aedes aegypti* e *Aedes albopictus* coletadas em pneus, no período de novembro de 1997 a outubro de 1998, em Nova Iguaçu, Rio de Janeiro.

Número de larvas – *Ae.albopictus*

123	243	215	142	153	118	164	194	160	120
128	122	151	155	137	129	216	157	145	182

Número de larvas – *Ae.aegypti*

90	104	140	72	67	78	60	61	101	81
117	89	111	83	98	101	74	88	132	66

- Calcule a o número médio de larvas em cada grupo utilizando a média aritmética
- Calcule a o número médio de larvas em cada grupo utilizando a média geométrica
- Calcule o número mediano de larvas em cada grupo.
- Desenhe o *box plot* do número de larvas representando os dois grupos em um só gráfico.
- Comente o gráfico *box plot* quanto a dispersão dos dados, existência de valores aberrantes e simetria dos dados.

Atenção: é necessário ordenar os valores para fazer os itens c e d.

Exercício 34

Os dados a seguir são adaptados de estudo, publicado por Devicari et al. 2013, que avaliou o tamanho das asas (mm) de *Aedes scapularis* para machos e fêmeas da espécie, capturados no Município de São Paulo.

Machos

1,78	1,91	2,02	2,11	2,13	2,21	2,30	2,41	1,87	2,01
2,03	2,11	2,15	2,21	2,31	3,50	1,90	2,01	2,10	2,11
2,15	2,21	2,32							

Fêmeas

1,01	1,62	2,30	2,40	2,56	2,61	2,71	2,80	1,52	1,89
2,31	2,45	2,60	2,65	2,75	2,89	1,58	1,97	2,34	2,45
2,60	2,70	2,78							

- Calcule o tamanho médio da asa (mm) para cada sexo. Utilize a média aritmética;
- Calcule o tamanho mediano da asa (mm) para cada sexo;
- Calcule a variância, o desvio-padrão e o coeficiente de variação de Pearson do tamanho da asa (mm) para cada sexo.
- Machos e fêmeas são parecidos quanto ao tamanho da asa (mm)?
- E quanto à variabilidade?
- Apresente o box plot do tamanho da asa (mm) para os sexos e interprete o gráfico.

Exercício 35

A tabela abaixo foi extraída do artigo: Influência da Altitude, Latitude e Estação de Coleta (Regra de Bergmann) na dimensão de *Lutzomyia intermedia* (Lutz & Neiva, 1912) (Diptera:Psychodidae, Phlebotominae). Marcondes CB et al. (Memórias do Instituto Oswaldo Cruz, 1999;. vol94(5):693-700).

Discuta os resultados obtidos.

Dimensions (in μm) of females and some of their respective ratios for *Lutzomyia intermedia* from Viana, a low altitude and lower latitude locality in the State of Espírito Santo (ES), and from low altitude and higher latitude localities in the states of Rio de Janeiro and São Paulo

Structures and ratios	Viana (ES)				Rio de Janeiro and São Paulo			
	Mean	s	N	C. V.	Mean	s	N	C. V.
Width of head ^b	342.6	12.2	17	3.6	374.1	21.3	61	5.7
Length of eye ^b	200.4	11	17	5.5	222.1	14.2	56	6.4
Width of eye ^b	110.6	8.8	17	7.9	126.7	8.3	55	6.6
Length of palpomere 3 ^a	163.5	6.6	15	4	171.6	7.1	60	7.1
Length of palpomere 5 ^b	135.3	11.5	14	8.5	145.9	12.3	58	8.5
Total length of palpus ^b	552	21.5	14	3.9	580	29.3	58	5
Maximum width of wing ^b	576	29.9	13	5.2	620	36.4	64	5.9
Length wing/maximum width of wing ^a	3.53	0.201	13	5.7	3.35	0.225	61	6.7
Length of R ₂ ^a	547	30.4	13	5.5	585	58.9	65	9.7
δ^a	278	34.2	12	12.3	317	58.6	65	18.5
Length of R ₃ ^a	683	30.1	13	4.41	723	60.2	65	8.33
Length of anterior femur ^a	692	33.8	11	4.9	722.5	45.5	50	6.3
Maximum width of spermathecal head ^a	10.49	1.8	18	17.5	11.87	2.13	59	17.9

s: standard deviation; N: number of observations; C.V.: coefficient of variation; a: significant at 5%; b: significant at 1%; δ : distance between the distal extremity of R₁ and the fork of R₂₊₃.

Exercício 36

A tabela abaixo foi extraída do artigo: Influência da Altitude, Latitude e Estação de Coleta (Regra de Bergmann) na dimensão de *Lutzomyia intermedia* (Lutz & Neiva, 1912) (Diptera:Psychodidae, Phlebotominae). Marcondes CB et al. (Memórias do Instituto Oswaldo Cruz, 1999;. vol94(5):693-700).

Discuta os resultados obtidos.

Tabela - Distribuição das médias, desvios padrão e variâncias das absorbâncias de amostras de *Lutzomyia longipalpis* coletadas no campo, alimentadas em laboratório e de *Lutzomyia almerioi* procedentes de campo no período de 2002 a 2004.

Amostras/Ano	Local	<i>Lutzomyia longipalpis</i>			<i>Lutzomyia almerioi</i>		
		X	S	V	X	S	V
2002	campo	0,518	0,247	0,061	0,781	0,167	0,028
	laboratório	0,815	0,030	0,001	-	-	-
2003	campo	0,652	0,148	0,022	0,764	0,031	0,001
2004	campo	0,668	0,197	0,039	-	-	-

X = média; S = desvio-padrão; V = variância

Exercício 37

Os dados a seguir são provenientes de um estudo que avaliou o período de vida, em dias de 53 exemplares, machos e fêmeas, da espécie de *Triatoma sordida*, nos estágios de ninfa e adulto. Do id 1 ao 22 estão registrados machos da espécie no estágio de ninfa, id 23 ao id 53 registra fêmeas no estágio de ninfa, id 54 ao id 75, registra machos no estágio adulto, id 76 ao id 106 registra fêmeas no estágio de ninfa.

id	Tempo de vida (dias)	id	Tempo de vida (dias)	id	Tempo de vida (dias)	id	Tempo de vida (dias)
1	136	28	128	55	95	82	556
2	157	29	131	56	140	83	535
3	154	30	126	57	179	84	582
4	129	31	132	58	65	85	648
5	247	32	136	59	157	86	567
6	164	33	146	60	189	87	775
7	133	34	128	61	232	88	310
8	126	35	150	62	96	89	43
9	247	36	136	63	253	90	220
10	139	37	158	64	253	91	591
11	139	38	134	65	326	92	671
12	148	39	126	66	254	93	443
13	221	40	128	67	241	94	400
14	248	41	128	68	385	95	782
15	131	42	139	69	426	96	672
16	139	43	203	70	520	97	146
17	135	44	208	71	512	98	431
18	143	45	242	72	409	99	579
19	249	46	241	73	492	100	631
20	173	47	250	74	424	101	719
21	241	48	244	75	547	102	517
22	241	49	259	76	714	103	427
23	126	50	241	77	466	104	406
24	126	51	253	78	447	105	424
25	127	52	254	79	112	106	425
26	130	53	250	80	728		
27	129	54	253	81	625		

Fonte: Souza JMP de, 1987. *Triatoma sordida* – Considerações sobre o tempo de vida das formas adultas e sobre a oviposição das fêmeas. Revista de Saúde Pública. São Paulo, 12:291-6 (Adaptado).

- Apresente o tempo de vida (dias) de *T. sordida* em estágio de ninfa segundo sexo em um *box plot*;
- Apresente o tempo de vida (dias) de *T. sordida* em estágio de adulto segundo sexo em um *box plot*;
- Comente os gráficos quanto a dispersão dos dados, existência de valores aberrantes e simetria.

Tópicos iniciais de amostragem

População:

Em ecologia, a definição da população é necessária sempre que for de interesse descrever um grupo de indivíduos de uma espécie que está sob investigação. Para animais de grande porte, tal como baleias, elefantes, é possível expressar o tamanho da população pelo número de indivíduos. Para grande número de animais a exemplo dos insetos, o número de indivíduos precisa ser expresso por densidades por unidade de área ou volume ou por unidade de *habitat* que pode aumentar ou diminuir no tempo e no espaço. Estas modificações dependem da natalidade, mortalidade e deslocamento para fora e para dentro da mesma (dispersão).

Assim, o conceito de indivíduo é importante pois a população compreende indivíduos da mesma espécie. Para os organismos unitários (que apresentam desenvolvimento ontogenético previsível e determinado), caso da entomologia, a definição de população não é tão difícil. Ex: população de larvas de *Cx. quinquefasciatus* nas margens do Rio Pinheiros, Cidade de São Paulo

No estudo de insetos, a estimativa do número total de indivíduos pode ser feita por técnicas de estimação como a de captura e recaptura que envolve a captura e marcação (com tinta ou pó fluorescente) de um número de indivíduos (de uma espécie) e posterior soltura. Em um momento posterior, a captura de indivíduos marcados fornece uma estimativa do tamanho da população. Por exemplo capturam-se 100 indivíduos de uma espécie, que são marcados e devolvidos à população de origem. Se, na amostra posterior, de 100 indivíduos forem encontrados 50 com a marca, poderíamos considerar que, se a amostra é representativa da população, a proporção de indivíduos marcados num segundo momento reflete a mesma proporção encontrada no primeiro momento (onde foram marcados 100 indivíduos). Então, metade dos indivíduos foi marcada, sendo a população total composta por 200 indivíduos.

Técnica simplificada (porque não inclui mudanças na população, como a mortalidade, por exemplo)

De forma geral:

- a) Toma-se uma primeira amostra (n_1 indivíduos) de uma população com tamanho desconhecido (N) e marca-se estes indivíduos (r indivíduos marcados)
- b) Os indivíduos marcados são soltos e estes se misturam à população N
- c) Faz-se uma segunda amostragem de n_2 indivíduos e observa-se a quantidade de indivíduos marcados (m de um total de n_2 capturados).

Esta proporção deverá ser a mesma da existente na primeira amostragem (r de um total de N). Com isso, N pode ser estimado

Em números:

- a) Supor $n_1=300$;

- b) Indivíduos marcados ($r=300$) são soltos na população N
- c) Segunda amostragem $n_2=300$ e observa-se que, destes $r=60$ estão marcados (20%). Isto quer dizer que 80% não estão marcados
- d) Então a estimativa da população é $N= 1500$, que corresponde a ($N = \frac{n_1}{0,2} = \frac{300}{0,2} = 1500$)

Definições importantes

População absoluta: é o número de animais por unidade de área

Intensidade da população: número de animais por unidade de *habitat*, por exemplo, pneu, internódio de bambu. Também pode ser expresso pelo número de indivíduos por unidade de habitat/área. Considerar o esquema abaixo

Área de $10m^2$

<div> <div>pneu 1 0 larvas</div> <div>pneu 2 80 larvas</div> <div>pneu 3 100 larvas</div> </div>				
internódio bambu 1		internódio bambu 2	...	internódio bambu 10
10 larvas	...			Total 300 larvas

Intensidade da população por *habitat*

$$\text{Intensidade da população para pneus} = \frac{180}{3} = 60 \text{ larvas/pneu}$$

$$\text{Intensidade da população para internódio de bambu} = \frac{300}{10} = 30 \text{ larvas/internódio}$$

Avaliando por unidade de *habitat* por área

$$\text{Intensidade da população para pneus por } m^2 = \frac{180/3}{10} = 6 \text{ larvas por pneu por } m^2$$

$$\text{Intensidade da população para internódio de bambu por } m^2 = \frac{300/10}{10} = 3 \text{ larvas por internódio por } m^2$$

Riqueza: é a quantidade de espécies

Abundância: é a quantidade de indivíduos de uma espécie

Estimadores relativos:

São estimadores usados quando o número de capturas ou observações não podem ser expressos por densidade ou por unidade de área ou por habitat. Só permitem comparações no espaço e no tempo. Por exemplo a utilização de ovitrampa para identificar presença de *A. aegypti* numa área. Deve-se realizar a contagem de larvas e comparar no tempo e no espaço. Segundo Southwood (2000), com ajustes, esta estimativa pode ser utilizada como densidade.

População índice: quando o animal não é contado mas sim o seu produto. Ex: exuvia da larva

Elementos de análise: são unidades de observação. Podem ser domicílio, pessoa, mosquito, pneus, armadilhas em geral, plantas como as bromélias, oco das árvores, ou qualquer outra unidade.

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.

Por exemplo:

Parâmetro – proporção populacional (π)

Proporção de indivíduos da espécie *Haemagogus janthinomys/capricornii* em ocos de árvores da mata do Parque Estadual da Serra da Cantareira.

Parâmetro – média populacional (μ)

Quantidade média de larvas da espécie *Haemagogus janthinomys/capricornii* em ocos de árvores da mata do Parque Estadual da Serra da Cantareira.

Estimador: é uma fórmula matemática que permite calcular um valor (estimador por ponto) ou com um conjunto de valores (estimador por intervalo) para um parâmetro.

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

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

Proporção $\pi = \frac{m}{N}$ em que

m é a quantidade populacional de indivíduos da espécie *Haemagogus janthinomys/capricornii*
 N é a quantidade de indivíduos de todas as espécies

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

Indicações para utilizar uma amostra
População muito grande
Processo destrutivo de investigação

Vantagens de realizar um estudo com amostragem:
Menor custo
Menor tempo para obtenção dos resultados
Possibilidade de objetivos mais amplos
Dados possivelmente mais fidedignos

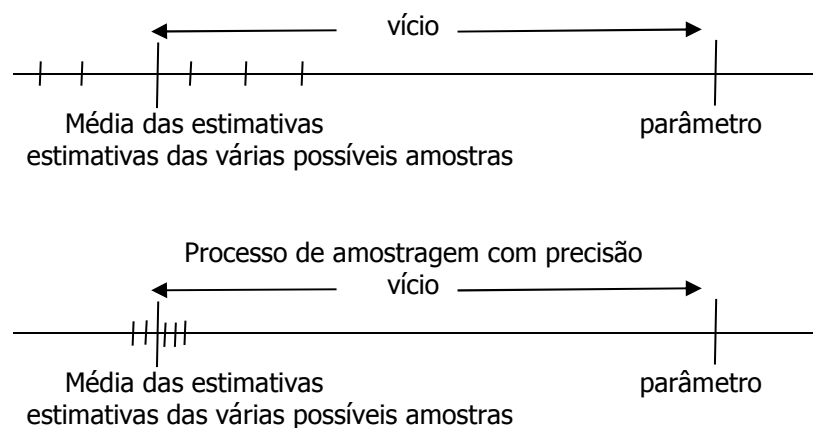
Desvantagens

Resultados por amostragem ficam sujeitos à variabilidade que pode ser entendida se for considerado que o processo de amostragem poderia fornecer um número muito grande de amostras. Cada amostra permite estimar o parâmetro. Então, os valores estimados por meio das várias amostras fornecem estimativas que flutam em torno da verdadeira média (parâmetro).

Erro e precisão

A distância entre a média das estimativas e o parâmetro constitui no erro de amostragem. Este erro não é observável. Mas pode ser diminuído melhorando o processo de coleta dos dados. Na prática se retira uma só amostra mas deve-se lembrar que ela é proveniente de uma população de amostras. O erro de amostragem também é chamado de vício, viés ou tendenciosidade.

A precisão está relacionada dispersão das estimativas em relação ao parâmetro podendo-se calcular o desvio padrão das médias. A precisão do processo de amostragem é dada pelo inverso do desvio padrão das médias. Então quanto menor a dispersão das estimativas em torno da média populacional, melhor a precisão do processo de amostragem. Esta precisão também é teórica sendo possível aumentá-la, trabalhando-se com amostras maiores.



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.

Supor que se deseja investigar quais são as espécies vetoras de febre amarela que ocorrem em ocos de árvores. As unidades de observação são árvores que apresentam ocos com constituição específica. A amostra será uma parte de todas as árvores contendo ocos de uma população de árvores de um determinado parque. O planejamento amostral define como serão identificados os ocos que serão estudados. Por exemplo pode-se decidir fazer uma amostragem por sorteio (probabilística) ou intencional (não probabilística)

Tipos de amostragem probabilística:

- aleatória simples (com e sem reposição);
- sistemática

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 (NN, 1998). 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.

Exercício 38

Apresente os passos de um plano de amostragem (probabilística e não probabilística) para a situação de espécies vetoras de febre amarela no parque.

Exercício 39

Os dados a seguir são de comprimento da asa de mosquitos Culicidae (mm) g).

X: Comprimento da asa de mosquitos Culicidae (mm)

2,30	2,34	2,40	2,47	2,54	2,60	2,64	2,64	2,68	2,71
2,74	2,74	2,74	2,74	2,74	2,76	2,76	2,76	2,79	2,79
2,82	2,83	2,84	2,84	2,84	2,89	2,90	2,92	2,93	2,93
2,93	2,93	2,93	2,94	2,94	2,94	2,94	2,94	2,94	2,94
2,99	3,00	3,02	3,02	3,02	3,02	3,02	3,02	3,04	3,04
3,04	3,04	3,04	3,04	3,04	3,04	3,04	3,04	3,05	3,05
3,05	3,05	3,05	3,08	3,08	3,09	3,09	3,09	3,09	3,11
3,11	3,11	3,11	3,11	3,12	3,12	3,12	3,12	3,14	3,14
3,14	3,14	3,15	3,15	3,17	3,19	3,19	3,20	3,20	3,20
3,20	3,24	3,24	3,24	3,24	3,29	3,29	3,29	3,29	3,31
3,31	3,34	3,39	3,44						

- Sorteie uma amostra aleatória de tamanho 20 utilizando a tabela dos números equiprováveis. Antes de fazer o sorteio, crie uma variável de identificação do indivíduo.
- Apresente os valores do comprimento da asa dos indivíduos sorteados.
- Some os valores e divida pelo tamanho da amostra (número de valores).
- Este valor é o parâmetro, o estimador ou a estimativa do comprimento médio?

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:

		elemento
i	1,941	1
$i+k$	$1,941+2,084 = 4,025$	4
$i+2k$	$1,941+4,1680 = 6,109$	6
$i+3k$	$1,941+6,252 = 8,193$	8
.	.	.
.	.	.
.	.	.
$i+(n-1)k$	$1,941+318,852 = 320,793$	320

Exercício 40 – Utilize os dados do Exercício 39.

- Sorteie uma amostra sistemática de tamanho 20. Indique o intervalo de amostragem e o começo casual sorteado. Indique o número de identificação de cada elemento da amostra.
- Some os valores e divida pelo tamanho da amostra (número de valores).
- Compare com o comprimento médio obtido no exercício 24. Você esperaria o mesmo resultado? Justifique.
- Qual dos dois valores você diria que representa melhor o conjunto de dados? Justifique.

Exercício 41

Explicar o processo de amostragem utilizado pelo Ministério da Saúde utilizando a publicação "Levantamento rápido de Índices para Vigilância Entomológica do *Aedes aegypti* no Brasil. Metodologia para avaliação dos índices de Breteau e predial de recipientes. Brasília – Df, 2013"

Noções de probabilidade e distribuição Bernoulli e distribuição binomial

PROBABILIDADE (*probability, chance, likelihood*)

- É uma afirmação numérica sobre a possibilidade de que algum evento ocorra.
- Quantifica o grau de incerteza de eventos, variando de 0 (0%) a 1 (100%).
- Um evento impossível de ocorrer tem probabilidade 0 (zero).
- Um evento certo tem probabilidade 1 (um).
- Quando se joga uma moeda, não se sabe se vai sair cara. Mas sabe-se que a probabilidade de sair cara é $0,5 = 50\% = 1/2$.
- Dizer que a eficácia de uma vacina é de 70% corresponde a dizer que cada indivíduo vacinado tem probabilidade 0,7 de ficar imune.

Probabilidade em espaços finitos contáveis

Espaço amostral (S)

- É o conjunto de todos os resultados possíveis de um experimento.
- Supor o experimento lançar uma moeda: $S = \{\text{cara, coroa}\}$

Há dois pontos neste espaço amostral, sendo um favorável ao evento $A = \{\text{cara}\}$.

Definição clássica de probabilidade

$$P(A) = \frac{\text{numero de elementos de } A}{\text{numero de elementos de } S} = \frac{1}{2} = 0,5$$

$$\text{Exemplo: probabilidade de (ouros)} = \frac{13}{52} = \frac{1}{4}$$

Probabilidade de eventos mutuamente excludentes

- Diz-se que dois eventos são mutuamente excludentes (ou mutuamente exclusivos) quando não podem ocorrer simultaneamente.

Exemplo:

$A = \{\text{cara}\}$; $B = \{\text{coroa}\}$, no lançamento de uma moeda;

$A = \{\text{carta com naipe vermelho}\}$; $B = \{\text{carta com naipe preto}\}$, na retirada de uma carta de baralho.

Exemplo de eventos não mutuamente exclusivos

$A = \{\text{naipe vermelho}\}$; $B = \{\text{ás}\}$.

- A probabilidade da ocorrência de um evento A ou de um evento B é:
 $P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$$\text{Exemplo: } P(\text{naipe vermelho ou ás}) = P(\text{naipe vermelho}) + P(\text{ás}) - P(\text{naipe vermelho e ás}) = (26/52) + (4/52) - (2/52) = 28/52 = 0,538.$$

- A probabilidade da ocorrência simultânea de eventos mutuamente exclusivos é zero.
 $P(\text{cara e coroa}) = P(\text{cara} \cap \text{coroa}) = 0$, no lançamento de uma moeda.

- Se A e B forem mutuamente excludentes, $P(A \cap B) = 0$ e
 $P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B)$

Exemplo:

P(Face 2 ou Face 3) no lançamento de um dado

$$P(2 \text{ ou } 3) = P(2) + P(3) = 1/6 + 1/6 = 2/6 = 1/3.$$

$$P(\text{Resultado ímpar}) = P(1 \text{ ou } 3 \text{ ou } 5) = P(1) + P(3) + P(5) = 3/6 = 1/2.$$

Regra da adição: $P(A \text{ ou } B) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Probabilidade de eventos independentes

- Os eventos A e B são independentes quando o resultado de um não influi no resultado do outro.

Exemplo: no lançamento simultâneo de duas moedas, o resultado de uma não interfere no resultado da outra.

- A probabilidade da ocorrência de eventos independentes é o produto das probabilidades de cada evento.

$$P(A \text{ e } B) = P(A \cap B) = P(A) \times P(B)$$

- P(face 2 no primeiro dado e face 3 no segundo dado), no lançamento sequencial de dois dados = $P(2 \text{ e } 3) = P(2) \times P(3) = 1/6 \times 1/6 = 1/36 = 0,0278 = 2,78\%$.

Probabilidade condicional

A probabilidade condicional do evento A dado que ocorreu o evento B é

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \text{ para } P(B) \neq 0$$

Lê-se $P(A|B)$ como probabilidade de A dado B.

Exemplo:

Probabilidade de rei dado que ocorreu figura:

$$P(r|figura) = P(r \text{ e } figura) / P(figura) = 4/52 \div 12/52 = 4/12 = 1/3$$

- Probabilidade de rei, dado que ocorreu copas:

$$P(r|\heartsuit) = P(r \text{ e } \heartsuit) / P(\heartsuit) = 1/52 \div 13/52 = 1/13$$

Regra da multiplicação

$P(A \cap B) = P(A B) \times P(B)$

se A e B forem independentes, $P(A B) = P(A)$ e como consequência, $P(A \cap B) = P(A) \times P(B)$

Exemplo

Considerar uma amostra de fêmeas de *Ae. triseriatus* segundo tamanho do corpo e estado do ciclo gonotrófico distribuídos segundo a tabela abaixo.

Distribuição de fêmeas de *Ae. triseriatus* segundo tamanho do corpo e estado do ciclo gonotrófico. Wisconsin, Estados Unidos, 1985.

Tamanho do corpo (mm) (X)	Paridas		Não paridas		Total	
	n	%	n	%	n	%
Pequeno ($X < 2,90$)	23		25		48	
Médio ($2,90 \leq X \leq 3,65$)	127		113		240	
Grande ($X \geq 3,66$)	12		28		40	
Total	162		166		328	

Fonte: Landry SV, et.al., 1988. (Adaptado). Journal of the American Mosquito Control Association vol4, nº2.

Escolhe-se uma fêmea ao acaso, qual a probabilidade dela ser parida dado que tem tamanho pequeno do corpo?

$$P(\text{parida}|\text{tamanho pequeno}) = \frac{P(\text{parida} \cap \text{tamanho pequeno})}{P(\text{tamanho pequeno})}$$

$$P(\text{parida}|\text{tamanho pequeno}) = \frac{\frac{23}{328}}{\frac{48}{328}} = \frac{23}{48} = 0,479 \text{ ou } 47,9\%$$

Escolhe-se uma fêmea ao acaso, qual a probabilidade dela ser parida dado que tem tamanho grande do corpo?

$$P(\text{parida}|\text{tamanho grande}) = \frac{P(\text{parida} \cap \text{tamanho grande})}{P(\text{tamanho grande})}$$

$$P(\text{parida}|\text{tamanho grande}) = \frac{\frac{12}{328}}{\frac{40}{328}} = \frac{12}{40} = 0,30 \text{ ou } 30,0\%$$

Os eventos ciclo gonotrófico e tamanho do corpo são independentes?

Dois eventos são independentes quando a probabilidade conjunta é igual ao produto das probabilidade individuais

Para os eventos serem independentes é necessário que a afirmação abaixo seja verdadeira

$P(\text{ser parida e ter corpo pequeno}) = P(\text{ser parida}) \times P(\text{ter corpo pequeno})$

$$\frac{23}{328} \stackrel{?}{=} \frac{162}{328} \times \frac{48}{328}$$

$$0,0701 \neq 0,4939 \times 0,1463$$

0,0701 \neq 0,0723 portanto, os eventos não são independentes.

Definição frequentista de probabilidade:

n repetições do evento A; A ocorre m vezes, então a frequência relativa de A = $\frac{m}{n}$

Para n suficientemente grande, $\frac{m}{n} \cong P(A)$ ou seja, $\lim_{n \rightarrow \infty} \frac{m}{n} = P(A)$

Quando n cresce, $\frac{m}{n}$ tende a se estabilizar em torno de uma constante, P(A)

Variável aleatória discreta

Variável aleatória é qualquer função de número real, definida no espaço amostral sendo que existe associado a este número uma probabilidade de ocorrência.

Exemplo:

Numa coleta de pupas de *A. aegypti* existe interesse em verificar se um indivíduo específico é fêmea.

Então define-se a variável X : número de pupas fêmeas.

Se for considerado 1 indivíduo coletado, X pode assumir valores 0 ou 1 que é representado como $X : 0, 1$

X será zero ($x=0$) se o indivíduo não for fêmea e será 1 ($x=1$) se for fêmea.

Se for atribuído um valor de probabilidade para $X=1$, denominado p , automaticamente tem-se o valor para probabilidade de $X=0$, denominado $1-p$ ou q . A probabilidade de $p+q=1$ que é o evento certo.

Supor que $p=0,6$ então $q=0,4$.

Então nestas condições, a variável número de fêmeas constitui uma variável aleatória.

$X: 0, 1$	
0	$\longrightarrow p = 0,6$
1	$\longrightarrow q = 0,4$

É possível calcular a probabilidade da variável assumir cada valor x , ou seja, $P(X=x)$.

O conjunto de valores da variável aleatória e das probabilidades obtidas define uma **distribuição de probabilidades**.

Se X assume valores inteiros, a variável é denominada discreta. Se X assume valores no conjunto dos números reais, a variável é denominada contínua.

Distribuição de probabilidades

Modelo de probabilidade Bernoulli

Estrutura básica: duas possibilidades de resultado (sucesso e fracasso).

Exemplo:

Joga-se uma moeda uma vez. A moeda é equilibrada, ou seja, os lados possuem peso igual, não favorecendo nenhum dos lados, ao ser lançada.

Define-se como sucesso sair cara.

Define-se a variável aleatória X que assume valor 1 se ocorrer sucesso e 0 se ocorrer fracasso. X : 0,1

Parâmetro: probabilidade da variável assumir valor 1.

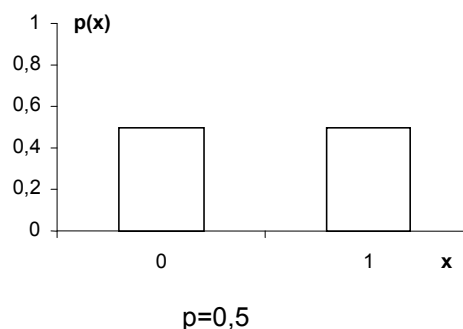
Notação: π ou p .

Se probabilidade de sucesso = p , a probabilidade de fracasso será igual a $q=(1-p)$, porque $p+q=1$.

Probabilidade de sair cara = $P(X=1) = p(1) = p = 0,5$.

Probabilidade de sair coroa = $P(X=0) = p(0) = q = 1-p = 0,5$

Graficamente:



Exemplo:

Estudos indicam que 27% de fêmeas de mosquitos Culicidae tem como fonte alimentar sangue de mamíferos. Realiza-se um estudo e captura-se indivíduos desta família. Ao observar um indivíduo, qual a probabilidade dele se alimentar desta fonte exclusivamente? Qual a probabilidade dele não se alimentar exclusivamente?

Sucesso: o mosquito se alimenta exclusivamente de sangue de mamíferos

X : número de indivíduos que se alimentam exclusivamente de sangue de mamíferos

X : número de sucessos

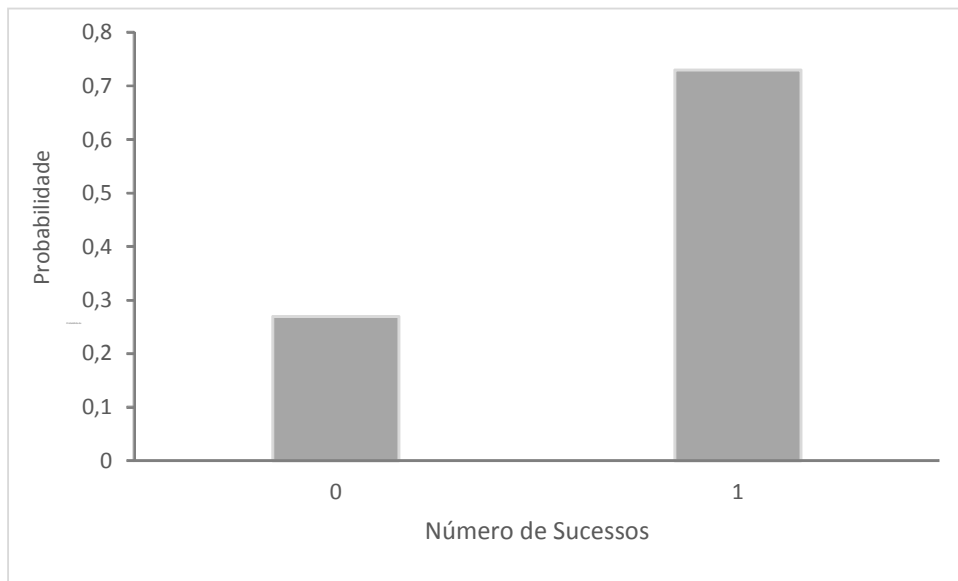
X : 0,1

(X será 0 se o mosquito alimentou-se de outra fonte que não exclusivamente de sangue de mamífero e 1 se alimentou-se exclusivamente de sangue de mamífero)

$P(X=1) = p(1) = p = 0,27$

$P(X=0) = p(0) = q = 0,73$

Representação gráfica



Forma geral da Distribuição de Bernoulli

$X: 0,1$

$$P(X=1) = p(1)=p$$

$$P(X=0) = p(0) = 1-p$$

$$\text{Então } P(x) = p^x (1-p)^{1-x}, x: 0,1$$

$$\text{para } X=0, p(0) = P(X=0) = p^0 (1-p)^{1-0} = 1-p,$$

$$\text{para } X=1, p(1) = P(X=1) = p^1 (1-p)^{1-1} = p$$

Média e Variância da Distribuição Bernoulli

De forma geral, a média de uma variável aleatória discreta: $\mu = E(X) = \sum_x xp(x)$

Na distribuição de Bernoulli:

$$\mu = E(X) = \sum_x xp(x) = 1p(x=1) + 0p(x=0) = p$$

Assim, a média da distribuição Bernoulli é p (probabilidade de ocorrer o sucesso)

De forma geral, a variância de uma variável aleatória discreta:

$$\sigma^2 = V(X) = E[(X - \mu)^2] = \sum_x (x - \mu)^2 p(x)$$

$$\text{Desvio padrão: } SD(X) = \sqrt{V(X)} = \sigma$$

Desvio padrão da distribuição Bernoulli é

$$\sqrt{(0-p)^2 \cdot p(x=0) + (1-p)^2 \cdot p(x=1)} =$$

$$\sqrt{(-p)^2 \cdot (1-p) + (1-p)^2 \cdot p} = \sqrt{p(1-p)[p + (1-p)]} = \sqrt{pq}$$

Resumindo,

Modelo de probabilidade Bernoulli

Uma variável aleatória discreta X que pode assumir valores 0 e 1, com função de probabilidade dada

$$\text{por } p(x) = p^x (1-p)^{1-x} \text{ com } x=0,1$$

X segue uma distribuição Bernoulli com parâmetro p , $0 < p < 1$.

p é a probabilidade de obter o resultado $X=1$. Isto pode ser escrito como $X \sim \text{Bernoulli}(p)$ com média p e desvio padrão $\sqrt{p(1-p)}$.

O símbolo \sim lê-se “tem distribuição” ou se “distribui segundo”.

Distribuição binomial:

Soma de n distribuições Bernoulli

População: 2 categorias

Ex: Sexo de pacientes (masculino, feminino),
Sexo dos insetos (macho, fêmea)
Fases de uma moeda (cara, coroa),
Desfecho de um tratamento (cura, não cura)

$$\text{Lançamento de uma moeda} \left\{ \begin{array}{l} \text{Cara (K)} \rightarrow \text{probabilidade(K)} = p \\ \text{Coroa (C)} \rightarrow \text{probabilidade (C)} = q \\ p + q = 1 \Rightarrow q = 1 - p \end{array} \right.$$

p = probabilidade de sucesso; q = probabilidade de fracasso.

Realiza-se o experimento n vezes, onde cada ensaio é independente do outro e os resultados são mutuamente exclusivos.

X : Número de sucessos

X : Número de vezes que sai cara.

A moeda é lançada uma vez ($n=1$) → $X: 0,1$ $X \sim \text{Bernoulli}(p)$

X	resultado	$P(X=x)$
0	C	$P(X=0) = q$
1	K	$P(X=1) = p$

A moeda é lançada duas vezes ($n=2$) → $X: 0,1,2$ $X \sim B(n=2, p)$

X	resultado	$P(X=x)$
0	C,C	$P(X=0) = q \cdot q = q^2$
1	K,C ou C,K	$P(X=1) = p \cdot q + q \cdot p = 2 \cdot p \cdot q$
2	K,K	$P(X=2) = p \cdot p = p^2$

A moeda é lançada três vezes ($n=3$) → $X: 0,1,2,3$ $X \sim B(n=3, p)$

X	resultado	$P(X=x)$
0	C,C,C	$P(X=0) = q \cdot q \cdot q = q^3$
1	K,C,C ou C,K,C ou C,C,K	$P(X=1) = p \cdot q \cdot q + q \cdot p \cdot q + q \cdot q \cdot p = 3 p \cdot q^2$
2	K,K,C ou K,C,K ou C,K,K	$P(X=2) = p \cdot p \cdot q + p \cdot q \cdot p + q \cdot p \cdot p = 3 p^2 \cdot q$
3	K,K,K	$P(X=3) = p \cdot p \cdot p = p^3$

Probabilidade ($X=x$) é calculada pelo produto de 3 fatores:

- 1º - número (combinação de n elementos combinados x a x);
- 2º - probabilidade de sucesso elevado a um expoente (valor de x);
- 3º - probabilidade de fracasso elevado a um expoente (valor de $n-x$).

$$P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

Resumindo

Modelo de probabilidade Binomial

Seja E um experimento com 2 resultados (mutuamente exclusivos): S (sucesso) e F (fracasso)

p = probabilidade de ocorrência de sucesso e q = probabilidade de ocorrência de fracasso sendo que $p+q=1$.

Se E for repetido n vezes, de forma independente, mantendo-se p e q constantes, a probabilidade da variável aleatória X = número de vezes que S ocorre é dada por

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$X \sim B(n,p)$ onde n e p são os parâmetros da distribuição; a média = $\mu = n \cdot p$, a variância = $\sigma^2 = n \cdot p \cdot q$

e o desvio padrão = $\sigma = \sqrt{npq}$

Exemplo

Lançamento de moedas.

n = número de ensaios (n° de lançamentos) = 10;

X = variável aleatória (n° de caras);

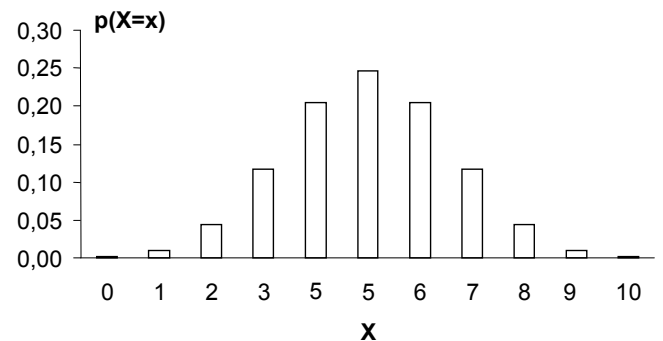
x = resultado particular de X (0, 1, 2, ..., 10);

p = probabilidade de ocorrer cara (sucesso); $p = P(\text{cara}) = 0,5$.

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

Distribuição de probabilidade $B(n=10; p=0,5)$

$X = n^\circ$ de caras	$P(X=x)$
0	0,0010
1	0,0098
2	0,0439
3	0,1172
4	0,2051
5	0,2461
6	0,2051
7	0,1172
8	0,0439
9	0,0098
10	0,0010
1	



Média = $np = 10 \times 0,5 = 5$.

Variância = $npq = 2,5$.

Desvio padrão = $\sqrt{npq} = \sqrt{10 \times 0,5 \times 0,5} = \sqrt{2,5} = 1,58$.

Se estivermos trabalhando com a proporção de sucessos, $\frac{X}{n}$:

$$\text{Média} = n \frac{p}{n} = p = 0,5$$

$$\text{Variância} = n \frac{p}{n} \times \frac{q}{n} = \frac{pq}{n} = 0,025$$

$$\text{Desvio padrão} = \frac{\sqrt{npq}}{n} = \sqrt{\frac{npq}{n^2}} = \sqrt{\frac{pq}{n}} = 0,158$$

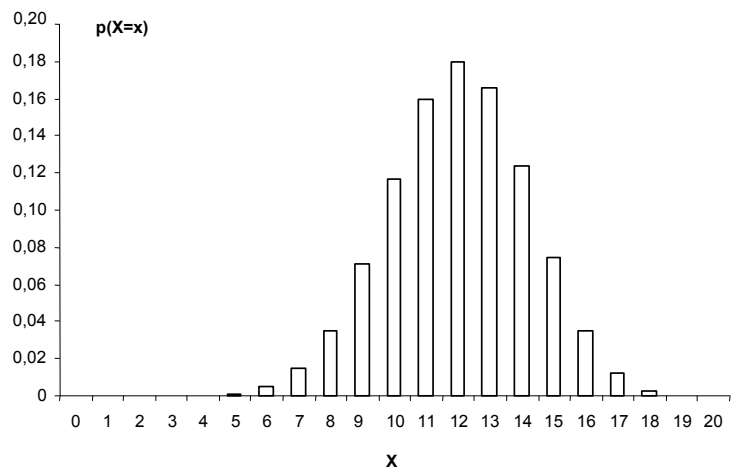
Exemplo

Observa-se em laboratório, fêmeas de *A. aegypti* com o objetivo de estudar o ciclo gonotrófico. Em uma amostra de 20 fêmeas definiu-se como sucesso, a ocorrência de oviposição nas 72 primeiras horas. Sabendo-se que a probabilidade de ovipor é de 60%, construa a distribuição de probabilidades para a variável X: número de sucessos.

X: Número de fêmeas que ovopõem nas 72 primeiras horas

X: 0, 1, 2, 3, ..., 20

X= nº de fêmeas que ovipõem	P(X=x p=0,6)
0	0,000
1	0,000
2	0,000
3	0,000
4	0,000
5	0,001
6	0,005
7	0,015
8	0,035
9	0,071
10	0,117
11	0,160
12	0,180
13	0,166
14	0,124
15	0,075
16	0,035
17	0,012
18	0,003
19	0,000
20	0,000



Calcule a média, a variância e o desvio-padrão.

Exemplo

Uma suspensão contendo organismos de *Leishmania* é preparada e quando uma determinada quantidade é inoculada em ratos, 30% deles se tornam infectados. Se 3 ratos forem inoculados independentemente, qual a probabilidade de:

a) Nenhum rato ficar infectado?

$$P(X=0) = \binom{3}{0} (0,3)^0 (0,7)^3 = \frac{3!}{0!(3-0)!} (0,3)^0 (0,7)^3 = 1 \times 0,343 = 0,343 = 34,3\%$$

b) Um rato ficar infectado?

$$P(X=1) = \binom{3}{1} (0,3)^1 (0,7)^{3-1} = \frac{3!}{1!(3-1)!} (0,3)^1 (0,7)^{3-1} = \frac{3 \times 2 \times 1}{1 \times 2 \times 1} 0,3 \times 0,49 = 0,441 = 44,1\%$$

c) Dois ratos ficarem infectado?

$$P(X=2) = \binom{3}{2} (0,3)^2 (0,7)^{3-2} = \frac{3!}{2!(3-2)!} (0,3)^2 (0,7)^{3-2} = \frac{3 \times 2 \times 1}{2 \times 1 \times 1} 0,09 \times 0,7 = 0,189 = 18,9\%$$

d) Todos os ratos ficarem infectados?

$$P(X=3) = \binom{3}{3} (0,3)^3 (0,7)^{3-3} = \frac{3!}{3!(3-3)!} (0,3)^3 (0,7)^0 = \frac{3 \times 2 \times 1}{3 \times 2 \times 1 \times 1} 0,027 \times 1 = 0,027 = 2,7\%$$

e) Pelo menos 2 fiquem infectados?

f) No máximo 1 fique infectado?

Exercício 42

Supor um teste com questões com 5 respostas de múltipla escolha com somente uma alternativa correta.

- a) Se o aluno escolher uma ao acaso ("chute"), qual a probabilidade dele escolher a resposta certa?
- b) Supondo que o teste tenha 20 questões; definindo-se a variável aleatória T: número de questões certas, qual é a distribuição de probabilidade da variável T?
- c) Calcular a probabilidade de um aluno acertar, no chute, 3 questões.
- d) Se o escore mínimo para passar é 10, qual a probabilidade de um aluno passar no teste, somente chutando?
- e) Qual o número médio de acertos esperado se o aluno somente chutar as respostas?

Exercício 43

Certa doença tem letalidade de 70%. Supondo-se que existam 20 pacientes com esta doença, calcular:

- a) a probabilidade de que todos morram da doença.
- b) a probabilidade de que nenhum paciente morra da doença.
- c) a probabilidade de que 7 pacientes morram da doença.
- d) a probabilidade de que, no máximo, 10 pacientes morram da doença.
- e) a probabilidade de que, no mínimo, 5 pacientes sobrevivam.
- f) o número esperado de óbitos e o respectivo desvio padrão.

Exercício 44

Em uma grande população, 20% das pessoas são canhotas. Assumindo que a variável X: número de pessoas canhotas segue uma distribuição Binomial, e sorteando-se uma amostra aleatória de 10 pessoas, encontre a probabilidade de:

- a) encontrar 2 pessoas canhotas .
- b) encontrar pelo menos 2 pessoas canhotas.
- c) encontrar no máximo 1 pessoa canhota.
- d) encontrar de 1 a 4 pessoas canhotas.

Exercício 45

Um caso de esquistossomíase é identificado pela detecção de ovo de shistosoma em amostra de fezes. Em pacientes com infecção baixa, uma técnica de exame de fezes tem probabilidade de 0,4 de detectar ovo. Se 5 amostras são examinadas para cada paciente, qual a probabilidade de um paciente com baixa infecção não ser identificado?

Exercício 46

São capturadas 15 fêmeas de *Anopheles darlingi* com o objetivo de investigar a infectividade por *Plasmodium vivax*. Sabendo-se que a probabilidade de infecção por *P. vivax* é de 0,10, calcule a probabilidade de

- a) Nenhuma fêmea estar infectada
- b) Todas as fêmeas estarem infectadas
- c) Duas fêmeas estarem infectadas
- d) Pelo menos 4 fêmeas estarem infectadas
- e) No máximo 2 fêmeas estarem infectadas
- f) Calcule o número médio de fêmeas infectadas e o desvio padrão

Exercício 47

Supor que 20% de coletas de adultos de *A. aegypti* no Parque do Carmo, em São Paulo resulte em machos. Uma coleta no parque ecológico do Tietê, resultou em 25 exemplares desta espécie. Calcule a probabilidade de que:

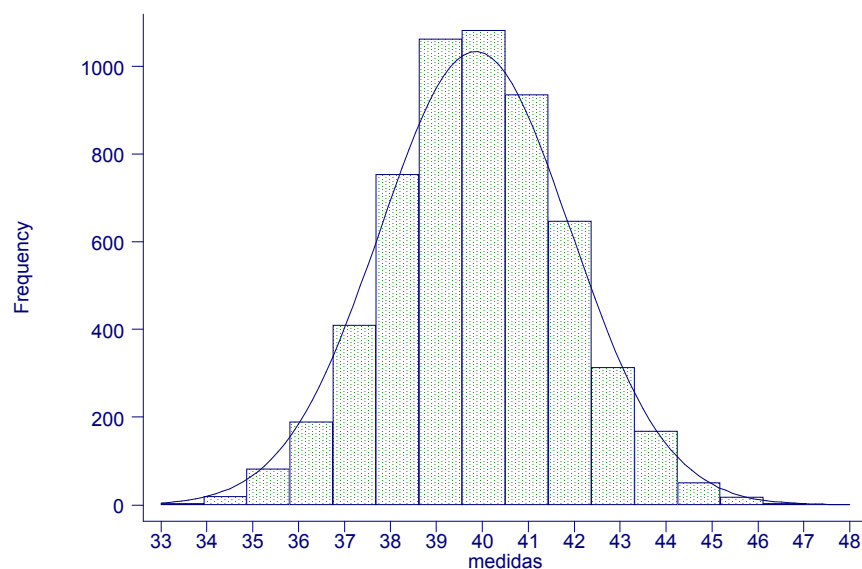
- a) Todos exemplares sejam fêmeas;
- b) Todos os exemplares sejam machos;
- c) 3 exemplares sejam fêmeas
- d) Pelo menos 5 exemplares sejam fêmeas;
- e) No máximo 5 exemplares sejam fêmeas;
- f) Qual é o número esperado de fêmeas e o desvio padrão?

Distribuição normal ou de Gauss; distribuição amostral da média

Os dados abaixo são medidas do tórax (polegadas) de 5732 soldados escoceses, tomadas pelo matemático belga, Adolphe Quetelet (1796-1874).

medidas	Freq,	Percent	Cum,
33	3	0,05	0,05
34	19	0,33	0,38
35	81	1,41	1,80
36	189	3,30	5,09
37	409	7,14	12,23
38	753	13,14	25,37
39	1062	18,53	43,89
40	1082	18,88	62,77
41	935	16,31	79,08
42	646	11,27	90,35
43	313	5,46	95,81
44	168	2,93	98,74
45	50	0,87	99,62
46	18	0,31	99,93
47	3	0,05	99,98
48	1	0,02	100,00
Total	5732	100,00	

Distribuição de medidas do tórax (polegadas) de soldados escoceses.



Fonte: Daly F et al. Elements of Statistics, 1999.

Função densidade de probabilidade da distribuição normal: Se a variável aleatória X é normalmente distribuída com média μ e desvio padrão σ (variância σ^2), a função densidade de probabilidade de

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]}$$

X é dada por

$$, -\infty < x < +\infty ;$$

onde

π : constante $\cong 3,1416$; e: constante $\cong 2,718$

μ : constante (média aritmética da população)

σ : constante (desvio padrão populacional)

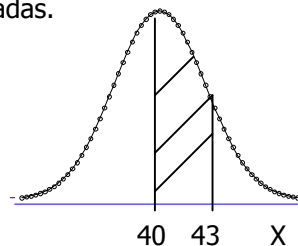
Propriedades:

- campo de variação : $-\infty < X < +\infty$;
- é simétrica em torno da média μ (ou m);
- a média e a mediana são coincidentes;
- a área total sob a curva é igual a 1 ou 100%;
- a área sob a curva pode ser entendida como medida de probabilidade.

$$\left\{ \begin{array}{l} \mu \pm 1\sigma \text{ inclui } 68,2\% \text{ das observações} \\ \mu \pm 1,96\sigma \text{ inclui } 95,0\% \text{ das observações} \\ \mu \pm 2,58\sigma \text{ inclui } 99,0\% \text{ das observações} \end{array} \right.$$

Exemplo:

Depois de tomarmos várias amostras, decidiu-se adotar um modelo para as medidas de perímetro do tórax de uma população de homens adultos com os parâmetros: média (μ) = 40 polegadas e desvio padrão (σ) = 2 polegadas.



Qual a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax entre 40 e 43 polegadas?

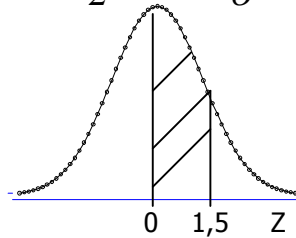
$$P(40 \leq X \leq 43) = \int_{40}^{43} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Quanto desvio padrão 43 está em torno da média?

Normal reduzida:

$$Z \sim N(0;1) \quad \text{onde } Z = \frac{X - \mu}{\sigma}$$

$$P(40 \leq X \leq 43) = P\left(\frac{40 - 40}{2} \leq \frac{X - \mu}{\sigma} \leq \frac{43 - 40}{2}\right) = P(0 \leq Z \leq 1,5)$$



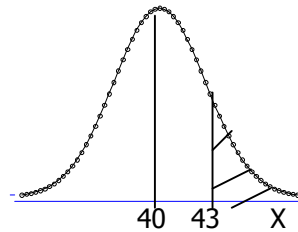
Utilizando a tabela da curva normal reduzida,

$$P(0 \leq Z \leq 1,5) = 0,43319 = 43,3\%$$

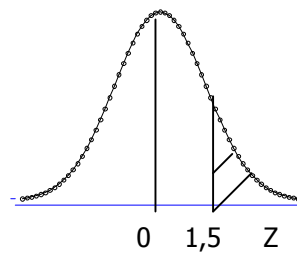
Exemplo

Com base na distribuição de $X \sim N(\mu = 40, \sigma = 2)$, calcular:

a) a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax maior ou igual a 43 polegadas.



$$P(X \geq 43) = P\left(\frac{X - \mu}{\sigma} \geq \frac{43 - 40}{2}\right) = P(Z \geq 1,5)$$



Utilizando a tabela da curva normal reduzida,

$$P(Z \geq 1,5) = 0,5 - 0,43319 = 0,06681 = 6,7\%.$$

b) a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax entre 35 e 40 polegadas.

c) a probabilidade de um indivíduo, sorteado desta população, ter um perímetro de tórax menor que 35.

d) Qual o valor do perímetro do tórax, que seria ultrapassado por 25% da população?

Exercício 48

Considere que fêmeas de *Anopheles darlingi*, criadas em laboratório apresentam peso seco (mg) com média 0,20 mg e desvio padrão 0,06 mg. Sorteia-se um exemplar; qual a probabilidade de que ele tenha

- a) Peso seco entre 0,18 e 0,21 mg?
- b) Peso seco maior que 0,27 mg
- c) Peso seco maior que 0,20 mg
- d) Peso seco menor que 0,15 mg
- e) Calcule o valor do peso seco que deixaria 25% da população de fêmeas abaixo dele.
- f) Calcule o valor do peso seco que deixaria 25% da população de fêmeas acima dele.

Exercício 49

Considere que fêmeas de *Aedes triseriatus*, apresentam comprimento da asa (mm) média 3,25 mm e desvio padrão 0,30 mm. Sorteia-se um exemplar; qual a probabilidade de que ele tenha

- a) Comprimento da asa entre 3,0 e 3,50 mm?
- b) Comprimento da asa maior que 3,75 mm
- c) Comprimento da asa maior que 3,00 mm
- d) Comprimento da asa menor que 3,55 mm
- e) Calcule o valor do comprimento da asa que deixaria 5% da população de fêmeas abaixo dele.
- f) Calcule o valor do comprimento da asa que deixaria 95% da população de fêmeas abaixo dele.

Distribuição amostral da média

Supor a situação onde uma população é composta por 6 elementos, para os quais observou-se a característica X, cujos valores estão apresentados abaixo.

elementos	X_i
A	11
B	16
C	12
D	15
E	16
F	14

Fonte: Dixon WJ e Massey FJ. Introduction to Statistical Analysis. 2nd edit. The Maple Press Company, York, 1957.

Média populacional (μ) = 14;

Variância populacional (σ^2) = 3,667;

Desvio padrão populacional (σ) = 1,9149.

Parâmetros População	valor	Estimador amostra	Valor (estimativa) Par(A,D)=(11,15)
Média (μ)	14	\bar{x}	13
Variância (σ^2)	3,67	S^2	8
Desvio padrão (σ)	1,91	S	2,828

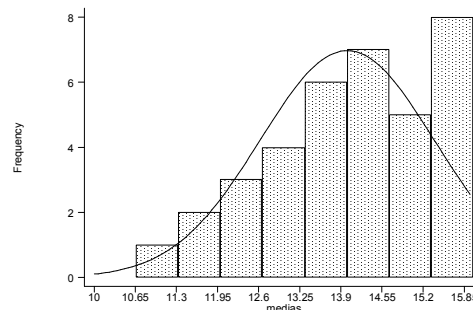
Todas as possíveis amostras de tamanho 2, determinadas pelo processo de amostragem aleatório, com reposição (N=6, n=2):

Amostra	Elementos que compõem a amostra	valores	Média(\bar{x}_i)
1	A,A	(11,11)	11
2	A,B	(11,16)	13,5
3	A,C	(11,12)	11,5
4	A,D	(11,15)	13
5	A,E	(11,16)	13,5
6	A,F	(11,14)	12,5
7	B,A	(16,11)	13,5
8	B,B	(16,16)	16
9	B,C	(16,12)	14
10	B,D	(16,15)	15,5
11	B,E	(16,16)	16
12	B,F	(16,14)	15
13	C,A	(12,11)	11,5
14	C,B	(12,16)	14
15	C,C	(12,12)	12
16	C,D	(12,15)	13,5
17	C,E	(12,16)	14
18	C,F	(12,14)	13
19	D,A	(15,11)	13
20	D,B	(15,16)	15,5
21	D,C	(15,12)	13,5
22	D,D	(15,15)	15
23	D,E	(15,16)	15,5
24	D,F	(15,14)	14,5
25	E,A	(16,11)	13,5
26	E,B	(16,16)	16
27	E,C	(16,12)	14
28	E,D	(16,15)	15,5
29	E,E	(16,16)	16
30	E,F	(16,14)	15
31	F,A	(14,11)	12,5
32	F,B	(14,16)	15
33	F,C	(14,12)	13
34	F,D	(14,15)	14,5
35	F,E	(14,16)	15
36	F,F	(14,14)	14

Distribuição de frequência de todas as possíveis médias:

Distribuição amostral da média

i	\bar{x}_i	frequência
1	11	1
2	11,5	2
3	12	1
4	12,5	2
5	13	4
6	13,5	6
7	14	5
8	14,5	2
9	15	5
10	15,5	4
11	16	4
Total		36



$$\text{Média das médias } (\bar{\bar{x}}) = \frac{\sum_{i=1}^{11} \bar{x}_i f_i}{n} = 14$$

$$\text{Variância das médias } \sigma_{\bar{x}}^2 = \frac{\sum_{i=1}^{11} (\bar{x}_i - \bar{\bar{x}})^2 f_i}{n} = 1,833 ;$$

$$\text{Desvio padrão das médias} = \text{erro padrão da média} = \sigma_{\bar{x}} = \sqrt{\sigma_{\bar{x}}^2} ;$$

$$\text{Erro padrão da média} = \sqrt{1,833} = 1,354 .$$

Teorema central do limite: X é variável aleatória com média μ e variância σ^2 , então

$$\bar{X} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

$$\text{No exemplo, } X \sim N(\mu = 14, \sigma = 1,915), \text{ portanto } \bar{X} \sim N(\mu_{\bar{x}} = 14, \sigma_{\bar{x}} = \frac{1,915}{\sqrt{2}} = 1,354) .$$

Exemplo

Os valores de ácido úrico em homens adultos sadios seguem distribuição aproximadamente Normal com média 5,7mg% e desvio padrão 1mg%. Encontre a probabilidade de que uma amostra aleatória de tamanho 9, sorteada desta população, tenha média

- maior do que 6 mg%.
- menor do que 5,2 mg%.

$$X \sim N(\mu = 5,7; \sigma = 1)$$

$$\begin{aligned} \text{a)} \quad P(\bar{X} \geq 6) &= P(Z_{\bar{X}} \geq \frac{6-5,7}{\frac{1}{\sqrt{9}}}) = P(Z_{\bar{X}} \geq 0,91) = 0,5 - 0,31859 = 0,18141. \\ \text{b)} \quad P(\bar{X} \leq 5,2) &= P(Z_{\bar{X}} \leq \frac{5,2-5,7}{\frac{1}{\sqrt{9}}}) = P(Z_{\bar{X}} \leq -1,52) = 0,5 - 0,43574 = 0,064. \end{aligned}$$

Exercício 50

Considere uma amostra de 25 fêmeas de *Anopheles darlingi*, capturadas na floresta Amazônica. Sabe-se que a população desta espécie apresenta peso seco (mg) com média 0,20 mg e desvio padrão 0,06 mg. Calcule a probabilidade de que a amostra apresente

- a) Peso seco médio maior que 0,22 mg
- b) Peso seco médio maior que 0,19 mg
- c) Peso seco médio menor que 0,195 mg
- d) Peso seco médio entre 0,18 e 0,21 mg?

Exercício 51

Considere uma amostra de 9 fêmeas de *Aedes triseriatus*, capturadas em ambiente silvestre nos Estados Unidos. A população desta espécie é descrita na literatura como apresentando comprimento da asa (mm) com média 3,25 mm e desvio padrão 0,30 mm. Calcule a probabilidade de que a amostra apresente comprimento médio

- a) Entre 3,0 e 3,3 mm
- b) Maior que 3,35
- c) Maior que 3,28
- d) Menor que 3,20
- e) Entre 3,0 e 3,3

Teste de hipóteses, teste de hipóteses

Estatística descritiva

Descreve eventos por meio de:

- tabelas
- gráficos
- razões e índices
- parâmetros típicos (medidas de posição e dispersão)

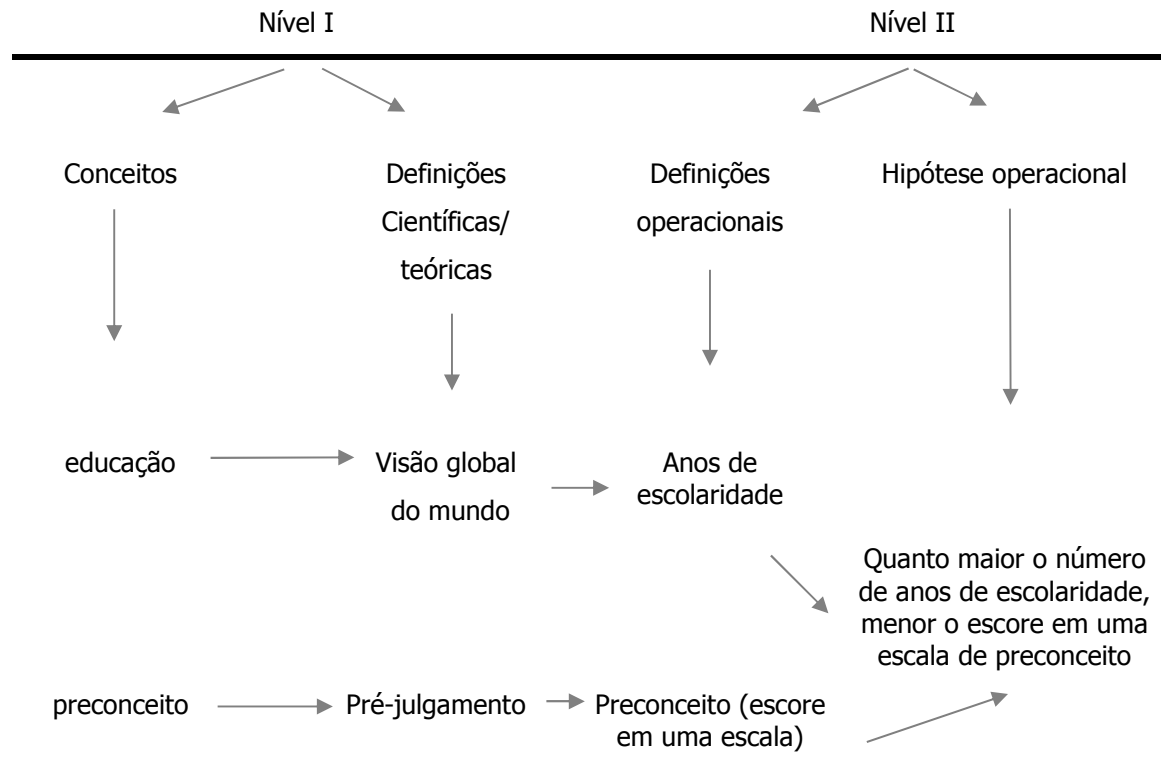
Estatística analítica

Nível I - Teórico (conceitos, hipóteses científicas)

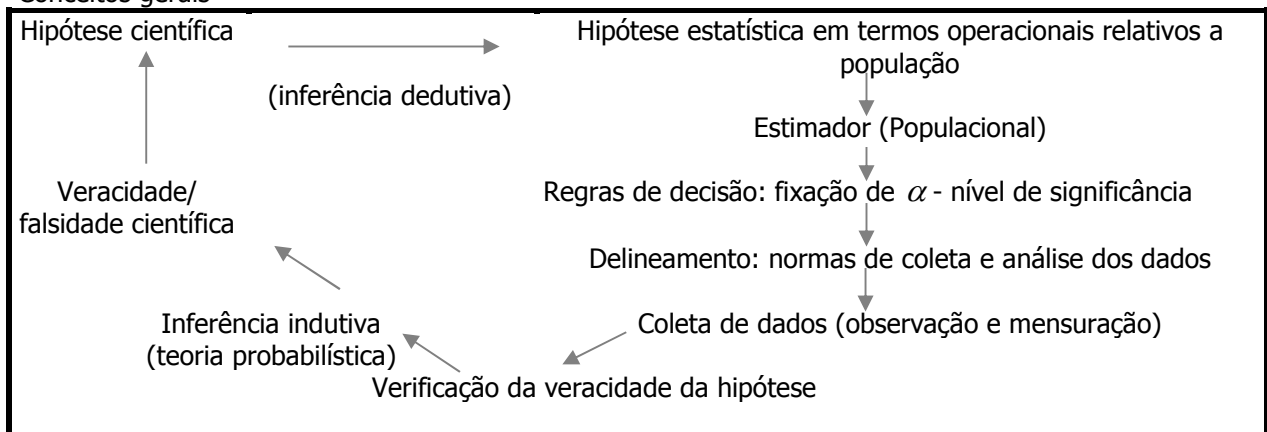
Nível II - operacional (hipótese estatística)

Situação

Quanto mais bem educada uma pessoa, menor o seu preconceito em aceitar certa campanha sanitária



Conceitos gerais



Inferência estatística: É qualquer procedimento que se utiliza para se generalizar afirmações sobre determinada população, baseadas em dados retirados de uma amostra.

Parâmetro: É a medida usada para se descrever uma característica de uma população.

Estatística: É uma função dos valores amostrais.

Estimação: É o processo através do qual estima-se o valor de um parâmetro de uma população com base no valor obtido em uma amostra.

Hipótese: É uma forma de especulação relativa a um fenômeno estudado (qualquer que seja). É qualquer afirmação sobre a distribuição de probabilidade de uma variável aleatória (afirmação sobre um parâmetro).

Hipótese estatística: É uma especulação feita em relação a uma proposição, porém relativa a uma população definida.

Teste de Hipóteses

Abordagem de Neyman e Pearson

Neyman e Pearson propuseram uma abordagem, para a tomada de decisão, que envolve a fixação, antes da realização do experimento, das hipóteses nula e alternativa, e fixação de valores de probabilidade de ocorrência de erros de decisão.

Considerar a situação na qual se deseja comparar a eficácia de uma nova droga (D_N) com a eficácia de uma droga padrão (D_A), que vem sendo atualmente utilizada.

Para a tomada de decisão sobre a eficácia de D_N , torna-se necessário seguir os seguintes passos:

- Formular as hipóteses;
- Identificar a distribuição de probabilidade da estatística do teste;
- Fixar o nível de significância do teste (α);
- Calcular o tamanho da amostra;
- Determinar a região de rejeição/aceitação de H_0 ;
- Realizar o estudo, observar os resultados, calcular a estatística do teste;
- Confrontar o valor observado da estatística do teste com a região de rejeição/aceitação de H_0 ;
- Tomar a decisão;
- Apresentar a conclusão.

Cada passo será apresentado detalhadamente a seguir.

Fixação das hipóteses para o exemplo da eficácia de D_N

Para o estudo proposto, onde uma nova droga é desenvolvida para apresentar maior eficácia que a droga em uso, as hipóteses apropriadas seriam:

$$\begin{cases} H_0 : D_N = D_A \\ H_a : D_N > D_A \end{cases} \quad \text{Teste monocaudal à direita}$$

Se o estudo envolvesse a comparação de duas drogas, uma nova e outra que é atualmente utilizada, e a nova droga se propõe a reduzir os efeitos colaterais, as hipóteses seriam:

$$\begin{cases} H_0 : D_N = D_A \\ H_a : D_N < D_A \end{cases} \quad \text{Teste monocaudal à esquerda}$$

Se ambas os lados forem possíveis, deve-se optar pela hipótese alternativa que explicita a diferença como na situação onde uma nova droga para depressão está em teste e deseja-se investigar se a

droga inibe ou provoca o apetite, como efeito colateral. Assim, antes do estudo não se conhece o efeito da droga sobre o apetite dos pacientes.

$$\begin{cases} H_0 : D_N = D_A \\ H_a : D_N \neq D_A \end{cases} \quad \text{Teste bicaudal}$$

Fixação de valores de probabilidade de ocorrência de erros de decisão

Considerar o estudo que tem por objetivo comparar a eficácia de uma nova droga (D_N) com a eficácia de uma droga padrão, que vem sendo utilizada (D_A), cuja eficácia é de 50%.

Eficácia (E) pode ser medida pelo número de curas. Isto indica qual é a estatística do teste que no caso é o número de curas.

Supor que a nova droga será utilizada em 10 pacientes ($n=10$) e, considerando-se a eficácia conhecida da droga antiga (D_A), de 50%, tem-se que a probabilidade de cura (p) é igual a 0,5.

$$\text{Hipóteses:} \quad \begin{cases} H_0 : D_N = 0,5 \\ H_a : D_N > 0,5 \end{cases} \quad \text{ou} \quad \begin{cases} H_0 : E_N = 0,5 \\ H_a : E_N > 0,5 \end{cases}$$

Estatística do teste: número de curas pela nova droga

X: número de curas, $X \sim B(n=10; p=0,50)$, se H_0 for verdade

X: 0, 1, 2, 3,...,10

Valor esperado de curas = $n.p = 10 \times 0,5 = 5$ curas

Distribuição de probabilidade Binomial para $n=10$ e $p=0,5$ (sob H_0 , ou seja, se H_0 for verdade)

X (número de curas)	P(X=x)
0	0,001
1	0,010
2	0,044
3	0,117
4	0,205
5	0,246
6	0,205
7	0,117
8	0,044
9	0,010
10	0,001

Utiliza-se o teste de hipóteses para testar H_0 . O teste de hipóteses fornece elementos para a tomada de decisão com base em H_0

É possível tomar somente uma decisão – Rejeita-se H_0 ou Não rejeita-se H_0 (Aceita-se H_0)

Possíveis erros na tomada da decisão:

Decisão	Verdade	
	H_0	H_a
H_0	não cometeu erro	<i>erro tipo II</i>
H_a	<i>erro tipo I</i>	não cometeu erro

$\alpha = \text{Pr obabilidade(erro tipo I)} = \text{Probabilidade (Rejeitar } H_0 \text{ e } H_0 \text{ é verdade)}$

$\beta = \text{Pr obabilidade(erro tipo II)} = \text{Probabilidade (Aceitar } H_0 \text{ e } H_0 \text{ é falsa)}$

$(1 - \beta) = \text{poder do teste} = \text{Probabilidade (Rejeitar } H_0 \text{ e } H_0 \text{ é falsa)}$

Poder de revelar a falsidade de H_0 quando a verdade é H_a

Conduta: Antes do experimento, fixa-se α e trabalha-se com o menor β possível.

Exercício 52

Em um julgamento jurídico o júri tem que decidir sobre a culpa ou inocência de um réu. Considere dois fatos: 1) o sistema jurídico admite que toda pessoa é inocente até que se prove o contrário. 2) só vai a julgamento pessoas sobre as quais existe dúvida de sua inocência. Fazendo uma analogia com teste de hipóteses, responda:

a) Apresente as hipóteses nula e alternativa sobre a culpa ou inocência do réu;

b) O júri pode errar se decidir que o réu é culpado quando na verdade ele é inocente; qual é o outro erro de decisão que o júri pode cometer?

c) Qual dos dois erros é o mais sério?

d) Na terminologia de teste de hipótese, qual tipo de erro (I ou II) pode-se vincular a cada uma das decisões do item b)?

Exercício 53

Supor duas situações: 1- a pessoa está fazendo parte de um levantamento para diagnóstico de para câncer de mama (*screening*); 2- a pessoa realiza o teste para detectar anticorpos anti-HIV.

É fornecido um diagnóstico com base no resultado do teste.

a) Qual dos erros é geralmente mais sério: um resultado falso positivo que diz que a pessoa tem a doença quando na verdade ela não tem ou um resultado falso negativo, que diz que a pessoa não tem a doença quando na verdade ela tem?

b) Apresente as hipóteses nula e alternativa sobre a situação de saúde do paciente; fazendo uma analogia com teste de hipóteses, que tipo de erro (I ou II) seria cometido se o resultado do teste fosse falso positivo? Que tipo de erro (I ou II) seria cometido se o resultado do teste fosse falso negativo?

Definição de critérios de aceitação ou rejeição de H_0 : estabelecimento das regiões de rejeição e de aceitação de H_0 .

Distribuição de probabilidade do número de curas sob H_0 : $B(n=10, p=0,5)$

X (número de curas)	P(X=x)	Região
0	0,001	Região de aceitação de H_0 $1-\alpha$
1	0,010	
2	0,044	
3	0,117	
4	0,205	
5	0,246	
6	0,205	
7	0,117	Região de rejeição de H_0 $\alpha = 5,5\%$
8	0,044	
9	0,010	
10	0,001	

Após a definição da área de rejeição de H_0 , pode-se realizar o experimento.

Por exemplo, supor que entre 10 pessoas que tomaram a nova droga, 9 se curaram. Como 9 cai na região de rejeição de H_0 , decide-se por rejeitar H_0 .

Se tivessem sido observadas 6 curas ou qualquer valor da área de aceitação de H_0 , a decisão seria não rejeitar H_0 ou seja, aceitar H_0 .

Onde está β ?

Lembrar que as hipóteses de teste são: $\begin{cases} H_0 : D_N = 0,5 \\ H_a : D_N > 0,5 \end{cases}$ e que a probabilidade do erro tipo II é a probabilidade de aceitar H_0 quando H_0 é falsa e que $(1 - \beta)$ é o poder do teste, ou seja, a probabilidade de rejeitar H_0 quando H_0 é falsa.

Supor que não se rejeita H_0 , portanto, decide-se por H_0 . Entretanto, se estiver sendo cometido algum erro de decisão, este será do tipo II. Assim, a verdade seria uma eficácia da nova droga maior que 0,5.

Supor que uma diferença de no mínimo 10% seja suficiente. Assim, supondo-se $p=0,6$, a distribuição do número de curas sob H_a , ou seja, sob uma $B(n=10, p=0,6)$ seria:

X (número de curas)	$p=0,5$ $P(X=x)$	Região	$p=0,6$	Região
0	0,001	$1 - \alpha$ aceitação de H_0	0,000	aceitação de H_0 $\beta = 0,833$
1	0,010		0,002	
2	0,044		0,011	
3	0,117		0,042	
4	0,205		0,111	
5	0,246		0,201	
6	0,205		0,251	
7	0,117	rejeição de H_0	0,215	rejeição de H_0
8	0,044		0,121	
9	0,010		0,040	
10	0,001		0,006	$(1 - \beta) = 0,167$

Notar que para n fixo, uma alteração no nível de significância, altera o poder do teste.

São apresentadas a seguir as relações entre o tamanho da amostra, o nível de significância, β e $1 - \beta$

Valores de β e de $1 - \beta$ para o teste de $H_0: E_N = E_P = 50\%$ contra $H_1: E_N > 50\%$, quando $n=10$, $\alpha \cong 5\%$ (a rigor, 5,47%) segundo diferentes valores de E_N .

E_N	β (%)	$1 - \beta$ (%)
60%	83,27	16,73
70%	61,72	38,28
80%	32,22	67,78
90%	7,02	92,98

Valores de β e de $1 - \beta$ para o teste de $H_0: E_N = E_P = 50\%$ contra $H_1: E_N > 50\%$, quando $n=10$, $\alpha \cong 1\%$ (a rigor, 1,08%) segundo diferentes valores de E_N .

E_N	β (%)	$1 - \beta$ (%)
60%	95,36	4,64
70%	85,07	14,93
80%	62,42	37,58
90%	26,39	73,61

Valores de β e de $1 - \beta$ para o teste de $H_0: E_N = E_P = 50\%$ contra $H_1: E_N = 60\%$, quando $\alpha \cong 5\%$ para diferentes valores de n .

Tamanho da amostra (n)	Valor de α mais próximo de 5%	Valor de β (%)	Valor de $1 - \beta$ (%)
10	5,5	83,3	16,7
15	5,9	78,3	21,7
20	5,7	75,0	25,0
25	5,4	72,6	27,4
30	4,9	70,9	29,1
35	4,5	69,4	30,6
40	4,0	68,3	31,7
45	6,8	67,3	32,7
50	5,9	55,4	44,6
55	5,2	54,1	44,9
60	4,6	54,9	45,1
65	4,1	54,7	45,3
70	6,0	48,8	51,2
75	5,3	45,0	55,0
80	4,6	45,2	54,8
100	4,4	37,7	62,3
150	4,3	22,6	77,4
160	4,8	18,7	81,3
175	4,8	15,8	84,2
200	5,2	11,0	89,0

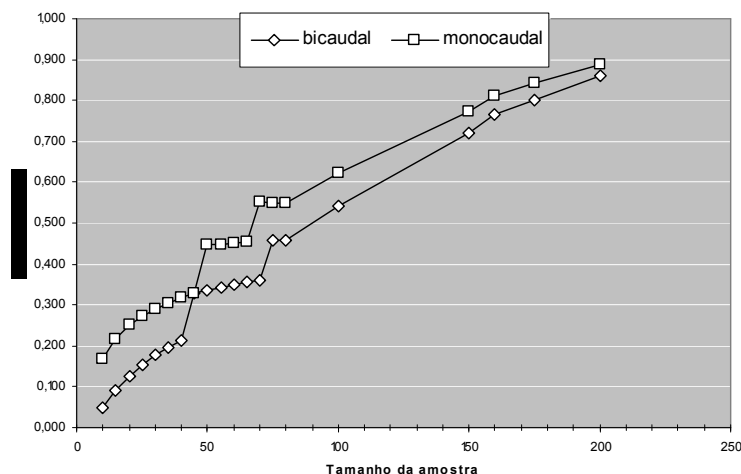
Valores de β e de $1 - \beta$ para o teste de $H_0: E_N = E_P = 50\%$ contra $H_1: E_N = 55\%$, quando $\alpha \cong 5\%$ para diferentes valores de n .

Tamanho da amostra (n)	Valor de α mais próximo de 5%	Valor de β (%)	Valor de $1 - \beta$ (%)
10	5,5	90,0	10,0
15	5,9	87,0	12,0
20	5,7	87,0	13,0
25	5,4	86,6	13,4
30	4,9	86,5	13,5
35	4,5	86,6	13,4
40	4,0	86,7	13,3
45	6,8	87,0	13,0
50	5,9	80,3	19,7
55	5,2	81,0	19,0
60	4,6	81,8	18,2
65	4,1	82,5	17,5
70	3,6	83,2	16,8
75	5,3	77,4	22,6
80	4,6	78,4	21,6
100	4,4	75,9	24,1
150	4,3	68,8	31,2
160	4,8	65,4	34,6
175	4,8	63,3	36,7
200	5,2	58,3	41,7
300	5,9	43,0	57,0
400	4,9	36,2	63,8
600	5,6	19,4	80,6

Para um teste bicaudal

Valores de β e de $1 - \beta$ para o teste de $H_0: E_N = E_P = 50\%$ contra $H_1: E_N \neq 50\%$, quando $\alpha \cong 5\%$ para diferentes valores de n e $E_1 = 60\%$ ou $E_1 = 40\%$

Tamanho da amostra (n)	Valor de α mais próximo de 5%	Valor de β (%)	Valor de $1 - \beta$ (%)
10	2,1	95,2	4,8
15	3,5	90,8	9,2
20	4,1	87,3	12,7
25	4,3	84,5	15,5
30	4,3	82,3	17,7
35	4,1	80,4	19,6
40	3,8	78,8	21,1
45	7,2	67,2	32,8
50	6,5	66,4	33,6
55	5,8	65,7	34,3
60	5,2	65,1	34,9
65	4,6	64,5	35,5
70	4,1	64,0	36,0
75	6,4	54,4	45,6
80	5,7	54,2	45,8
100	5,7	45,7	54,3
150	6,0	27,9	72,1
160	6,9	23,3	76,7
175	6,9	19,8	80,2
200	7,7	14,0	86,0



Poder do teste para tamanhos de amostra fixos em testes mono e bicaudal, com distribuições de probabilidade $B(n, p=0,5)$ para H_0 e $B(n, p=0,6)$ para H_a

Teste de hipóteses segundo a abordagem de Fisher (Ronald Aylmer Fisher)

Inicia-se a abordagem de Fisher com a especificação de uma proposição inicial (equivalente à H_0 de Neynman e Pearson). Considerando o estudo que tem por objetivo comparar a eficácia de uma nova droga (D_N) com a eficácia de uma droga padrão, que vem sendo utilizada (D_A), cuja eficácia é de 50%, tem-se:

Proposição inicial: $D_N=0,5$

Para tomada de decisão deve-se realizar o experimento e calcular a probabilidade de ocorrência do valor observado ou de um valor mais extremo da estatística do teste, em uma curva de probabilidade especificada na proposição inicial.

Se na amostra de 10 pacientes, 9 evoluíssem para a cura (90%), Fisher recomendava que se calculasse a probabilidade de 9 ou mais pacientes se curarem ($P(X \geq 9)$), tendo como base, a

distribuição de probabilidade conhecida, especificada na proposição inicial, onde a probabilidade de cura é igual a 50%.

Pelo exemplo, esta probabilidade seria igual a $P(X \geq 9) = P(X=9) + P(X=10) = 0,011 = 1,1\%$

Se na amostra de 10 pacientes, fossem observadas 6 curas (60%), $P(X \geq 6) = P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10) = 0,205 + 0,117 + 0,044 + 0,010 + 0,001 = 0,377 = 37,7\%$

A probabilidade calculada é conhecida como valor de p (*p-value*) e a decisão estatística será tomada com base no valor desta probabilidade.

Se o valor de p for considerado pequeno, conclui-se que os dados não mostram evidência de pertencer a uma população com proporção de cura igual a 50% e, portanto, a droga cura mais do que 50%.

Se o valor de p for considerado grande, então se pode dizer que os dados provavelmente vêm de uma população que possui como parâmetro 50% de curas.

Definição

Valor de p é a probabilidade de ocorrência do valor observado ou de um valor mais extremo de uma estatística, em uma curva de probabilidade especificada (conhecida, verdadeira).

Fisher dizia que antes de dar uma forma matemática a um problema, propondo hipóteses a serem testadas, era necessário um amplo conhecimento dos dados, o que poderia ser realizado com base no valor de p .

Passos necessários para a realização de um teste de hipóteses segundo a abordagem de Fisher.

- Formular a proposição inicial ("hipótese") que será testada;
- Identificar a distribuição de probabilidade;
- Realizar o estudo e observar o resultado da estatística de interesse;
- Calcular o valor de p , ou seja, a probabilidade de ocorrer o valor observado ou um valor mais extremo, sob a curva especificada na proposição inicial;
- Tomar a decisão com base no valor de p .
- Apresentar as conclusões

Teste de hipóteses para uma proporção populacional (Distribuição Binomial)

Considerar a situação

Segundo dados de rotina das ações de controle do *Aedes aegypti*, tem-se que, em determinada comunidade, a proporção de visitas domiciliares que resultam em identificação de domicílios **sem foco** para o vetor da Dengue é de 60%. Desejando-se aumentar esta proporção, realizou-se uma campanha educativa para envolver a população nas medidas de controle. Após um período de 6 meses o programa foi avaliado quanto a sua eficácia, observando-se que, em uma amostra de 10 domicílios desta comunidade em 9 não foram encontrados focos de *A. aegypti*.

Utilizando-se teste de hipóteses para decidir sobre a eficácia da intervenção:

Pela abordagem de Neyman e Pearson

- 1) Elaboração das hipóteses : $H_0 : E_p = 0,6$
 $H_a : E_p > 0,6$

- 2) Fixação de $\alpha = \text{Prob}(\text{rejeitar } H_0 \text{ e } H_0 \text{ é V})$; fixando-se $\alpha=0,05$

- 3) Estabelecimento da região de rejeição/aceitação de H_0 :
Estatística do teste: número de domicílios sem foco *A. aegypti*

X: 0,1,2,...,10

Eventos independentes e mutuamente exclusivos; portanto, a distribuição de probabilidade de X segue um modelo B(n=10; p=0,6)

Região de rejeição e aceitação de H_0 , se H_0 for verdade		
X	$P(X=x)$, sob H_0 p=0,6	Região
0	0,00010	$1 - \alpha$ Aceitação de H_0
1	0,00157	
2	0,01062	
3	0,04247	
4	0,11148	
5	0,20066	
6	0,25082	
7	0,21499	
8	0,12093	Rejeição de H_0 $\alpha = 0,04636 \cong 0,05$
9	0,04031	
10	0,00605	

- 4) Decisão: Como 9 domicílios livres de foco cai na área de rejeição de H_0 , decide-se por rejeitar H_0 .

5) Conclusão: Foi encontrada diferença estatisticamente significativa entre as proporções populacionais de domicílios sem foco antes e após o programa, para nível de significância de 5%. O programa educativo, portanto, foi eficaz, pois os domicílios da comunidade com pessoas que participaram do programa provêm de uma população de domicílios onde mais do que 60% estariam livres de foco do *A. aegypti*.

Abordagem de Fisher

Proposição: domicílios que são submetidos ao programa provêm de uma população onde 60% deles estão livres de foco de *A. aegypti*.

Calculando-se a probabilidade de observar 9 ou domicílios livres de foco de *A. aegypti*, utilizando-se uma curva onde 60% de domicílios estão livres de foco de *A. aegypti*:

Distribuição de probabilidade: $B(n=10; p=0,6)$, portanto

$$p = P(X \geq 9) = P(X = 9) + P(X = 10) = 0,040 + 0,006 = 0,046 \text{ ou } 4,6\%.$$

Interpretação do valor de p: 4,6% é a probabilidade de se observar 9 ou mais domicílios sem foco de *A. aegypti*, se estes tivessem vindo de uma população de domicílios onde 60% estariam sem foco de *A. aegypti*.

Para decidir com base no valor de p é necessário perguntar-se se os resultados observados são compatíveis com a proposição de domicílios que veem de população na qual 60% estão livres de foco de *A. aegypti*.

Com base nos resultados, você diria que existe evidência favorável ou contrária à proposição inicial?

Se p for considerada, pelo investigador, pequena então se conclui que os dados observados mostram evidência contrária à proposição inicial (a proporção de domicílios sem foco depois da campanha é mais compatível com uma população de domicílios onde mais do que 60% estariam sem foco de *A. aegypti*. Neste caso, poderia-se dizer que o programa foi eficaz.

Se p for considerado, pelo investigador, grande, então se conclui que os dados não mostram evidência contrária à proposição e, portanto, os domicílios, após a intervenção, devem ser de uma população na qual 60% não apresentam foco. Neste caso, pode-se inferir que a intervenção não surtiu efeito.

Considerações finais

O valor de p é a força de evidência contrária à proposição inicial. Para existir forte evidência contrária à proposição inicial, o valor de p deve ser bem pequeno;

O julgamento sobre o valor de p, se é grande ou pequeno, é arbitrário e quem decide é o investigador.

Exercício 54

Supor o experimento onde existe interesse em investigar se o odor em determinada armadilha atrai fêmeas de insetos hematófagos. O experimento consiste em colocar os espécimes em um corredor que no final é dividido para a direita e para a esquerda. Uma armadilha iscada com gelo seco, fonte de dióxido de carbono, é colocada no final do corredor à esquerda, fora da visão do inseto.

Entretanto, antes da realização do experimento, decide-se eliminar a possibilidade de incluir no estudo insetos que têm predileção por um lado, independentemente do odor da armadilha. Neste caso, decide-se investigar inicialmente se os insetos escolhem os lados em proporções iguais. Para tanto, realiza-se o experimento com 30 fêmeas de *A. aegypti* sem a colocação de armadilha com atrativo e verifica-se que 13 voam para a esquerda.

Realize um teste de hipóteses seguindo as propostas de Neyman e Pearson, com nível de significância de 5%, para verificar se os insetos vêm de uma população que escolhe mais um lado do que o outro.

Tome a decisão utilizando também a estratégia de Fisher

Exercício 55

Supor, agora, o experimento para investigar se o odor das armadilhas atrai fêmeas.

Realiza-se o experimento colocando-se a armadilha com atrativo no final do corredor do lado esquerdo, fora da visão dos insetos. Observa-se que de 30 fêmeas, 17 voam para a esquerda.

Realize um teste de hipóteses seguindo as propostas de Neyman e Pearson com nível de significância de 5% e a estratégia de Fisher, com cálculo do valor descritivo do teste, para verificar se as fêmeas vêm de uma população que escolhe mais o lado onde está o atrativo.

Exercício 56

Um estudo foi desenvolvido para investigar se uso de repelente protege pessoas contra picadas de insetos hematófagos após um período 6 horas. Considerando H_0 : repelente não protege após 6 horas da aplicação e H_a : repelente protege após 6 horas da aplicação, responda

- Tomando qual decisão sobre H_0 (aceitar ou rejeitar), você poderia estar cometendo o erro tipo I?
- Tomando qual decisão sobre H_0 (aceitar ou rejeitar), você poderia estar cometendo o erro tipo II?
- Como é denominada a probabilidade de ocorrência do erro tipo I?
- Como é denominada a probabilidade de ocorrência do erro tipo II?
- O que é o poder do teste?
- Se você fosse fixar valores de probabilidades associadas à ocorrência dos erros tipo I e II para este estudo, qual deles seria menor? Justifique.

Exercício 57

Será realizado um estudo para investigar a relação entre consumo de Vitamina B durante três meses anterior à exposição e ocorrência de picadas por insetos hematófagos. Para a tomada de decisão, será utilizado teste de hipóteses, no modelo clássico, proposto por Neyman e Pearson. Apresente os passos necessários para a realização do teste de hipóteses utilizando a listagem que segue com aplicação da situação apresentada.

Lista de passos: tomada de decisão, cálculo do tamanho da amostra, elaboração das hipóteses, determinação da região de rejeição do teste, coleta dos dados e cálculo da estatística do teste, fixação do nível de significância, conclusão, verificação se o valor observado da estatística cai na região de aceitação ou rejeição, identificação da distribuição de probabilidade da estatística do teste.

Exercício 58

Considere a seguinte situação hipotética: A incidência de dengue durante o verão, em uma município com infestação pelos vetores desta doença é 20%. Durante o ano de 2001, as autoridades sanitárias distribuíram mosquiteiros para berços de crianças para os moradores que se propuseram a usá-los. Após o verão observou-se que de 20 crianças que utilizaram os mosquiteiros, 1 contraiu dengue. Deseja-se saber se o uso de mosquiteiros teve efeito sobre a ocorrência de dengue.

Utilizando a abordagem de Fisher, responda:

- a) Qual seria a proposição inicial (equivalente à hipótese nula)?
- b) Se for definida a variável X: ter dengue, e considerando-se que foram acompanhadas 20 crianças, qual a distribuição de probabilidade de X? Especifique os parâmetros da distribuição.
- c) Fisher recomendava calcular o valor de p e decidir com base nele. O que é o valor de p?
- d) Calcule o valor de p e decida se os dados mostram evidência favorável ou desfavorável ao que for especificado na proposição inicial.
- e) Calcule o valor de p se 3 crianças tivessem ficado com dengue e decida sobre a propriedade de proteção do mosquiteiro.

Exercício 59

Utilize os dados do exercício anterior e responda as questões segundo a abordagem de Neyman e Pearson

- a) Qual seria a hipótese nula?
- b) Se for definida a variável X: ter dengue, e considerando-se que foram acompanhadas 20 crianças, qual a distribuição de probabilidade de X? Especifique os parâmetros da distribuição.
- c) É comum que se estabeleça os valores de probabilidade dos erros tipo I e tipo II antes de realizar o estudo. Como são denominadas estas probabilidades e que valores você fixaria para elas?

- d) Realize o teste de hipóteses para nível de significância de 2,5% e decida sobre a eficácia de utilização do mosquiteiro.
- e) Qual decisão você tomaria a respeito da eficácia do uso de mosquiteiros se 6 crianças tivessem contraído dengue?
- f) Qual decisão você tomaria se nenhuma criança tivesse contraído a doença?

Exercício 60

Em São Sebastião, litoral norte do estado de São Paulo observou-se em 2001 uma proporção de 76% de mosquitos *A. aegypti* no intradomicílio. Após intervenção ambiental com retirada de todos os potenciais criadouros em 20 domicílios, observou-se alteração desta proporção para 65%. Deseja-se saber se esta diferença foi devida ao acaso ou pode ser creditada a um efeito da intervenção. Utilize para tomada de decisão a abordagem clássica com nível de significância de 5%. Realize também o teste de hipóteses calculando o valor de p.

Exercício 61

A prevalência de infecção por hepatite B na população geral é de 30%. A literatura sugere que a infecção por hepatite B é maior entre pessoas com infectadas pelo vírus HIV. Em uma amostra de 20 pessoas que apresentaram teste HIV +, 8 apresentaram positividade para hepatite B. Teste a hipótese de que as pessoas HIV + possuem mesma prevalência de Hepatite B que a população geral. Utilize a estratégia clássica de Neyman e Pearson, com nível de significância de 5% e a abordagem de Fisher, com tomada de decisão a partir do valor descritivo do teste (valor de p).

Exercício 62

Dados de ações de controle da dengue em São Sebastião, litoral norte do estado de São Paulo, no período de fevereiro de 2011 a fevereiro de 2012, indicaram que entre 168 quarteirões investigados, 99 foram positivos para *A. aegypti*, com taxa de positividade igual a 58,9%. No primeiro semestre do período, de 28 quarteirões, 23 foram positivos. Realize um teste de hipóteses para investigar se a positividade de quarteirões no primeiro semestre é maior do que no período como um todo somente devido ao acaso. Utilize a estratégia clássica de Neyman e Pearson, com nível de significância de 5% e a abordagem de Fisher, com tomada de decisão a partir do valor descritivo do teste (valor de p).

Teste de hipóteses de uma média populacional (μ) com variância conhecida

Proposta clássica de Neyman e Pearson

Situação de interesse

Tomando-se como exemplo os dados de duração das fases de evolução (dias) de *Triatoma sordida* publicados, em condições de laboratório, Juarez E. Silva EPC (1982) descrevem que com temperaturas mais altas há um encurtamento do período como adulto. Considerando-se que a uma temperatura de 25°C, a duração na fase adulta segue distribuição normal com média populacional (μ) 224,8 dias e desvio padrão populacional (σ) 86,0 dias.

Para tanto foi realizado um estudo onde se observou o tempo médio na fase adulta de exemplares a uma temperatura de 30°C obtendo-se os tempos a seguir.

Deseja-se saber se existe efeito de temperatura na duração da fase adulta deste inseto

A variável de estudo X é tempo (dias) – variável resposta, desfecho ou outcome.

Tomando-se como referência a temperatura de 25°C, $X \sim N(\mu = 224,8; \sigma = 86,0)$.

Recordando-se, para a realização do teste de hipóteses segundo Neyman e Pearson é necessário:

- Formular as hipóteses estatísticas;
- Fixar a probabilidade do erro tipo I;
- Calcular o tamanho da amostra necessária para detectar uma diferença que se suspeita existente o que é equivalente a fixar a probabilidade do erro tipo II;
- Apresentar a distribuição de probabilidade da estatística do teste;
- Estabelecer a(s) região(ões) de rejeição e aceitação (regiões críticas) do teste;
- Realizar o estudo, ou seja, coletar os dados e calcular a estatística do teste;
- Confrontar a estatística do teste observada com a região crítica;
- Tomar a decisão;
- Elaborar a conclusão.

Formulação das hipóteses

$$\begin{array}{ll} H_0 : \mu_{30^\circ C} = \mu_{25^\circ C} & H_0 : \mu_{30^\circ C} = 224,8 \\ H_a : \mu_{30^\circ C} < \mu_{25^\circ C} & \text{ou} \quad H_a : \mu_{30^\circ C} < 224,8 \end{array}$$

Na situação de estudo, fixando-se o nível de significância $\alpha = 0,05$

Cálculo do tamanho mínimo da amostra

Para uma hipótese monocaudal, onde $H_0 : \mu_{30^\circ C} = 224,8$ e $H_a : \mu_{30^\circ C} < 224,8$

$$n = \frac{(Z_\alpha + Z_\beta)^2}{d^2}, \text{ em que}$$

Z_α é o valor de Z que deixa α à direita

Z_β é o valor de Z que deixa β à direita

$$d = \frac{|\mu_{30^\circ C} - \mu_{25^\circ C}|}{\sigma}$$

Supondo que a média populacional para $30^\circ C = 190$, $d = \frac{|190 - 224,8|}{86} = 0,4$

Pela tabela da $N(0,1)$ tem-se que para $\alpha = 0,05$, $Z_\alpha = 1,64$

Pela tabela da $N(0,1)$ tem-se que para $\beta = 0,20$, $Z_\beta = 0,845$

Substituindo-se os valores, tem-se

$$n = \frac{(Z_\alpha + Z_\beta)^2}{d^2} = \frac{(1,64 + 0,845)^2}{0,4^2} = 38,6$$

Exercício 63

Altere o valor de $\beta = 0,10$ e recalcule o tamanho da amostra.

Altere os valores de $\beta = 0,10$ e $d=0,3$ e recalcule o tamanho da amostra

Discuta os resultados

Coleta dos dados

Continuando sobre o teste de hipóteses, considerando-se que foi possível coletar dados para 22 exemplares de *Triatoma sordida*

136	157	154	135	247
164	140	126	247	139
139	148	221	248	150
139	135	143	249	173
241	241			

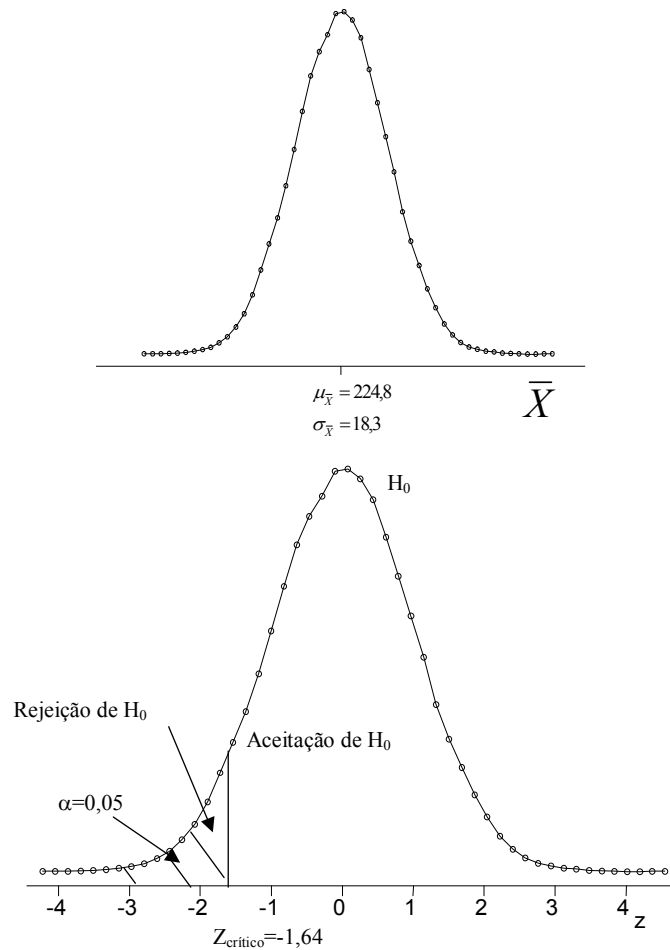
Distribuição de probabilidade

Como as hipóteses envolvem a média populacional, é necessário utilizar a distribuição de probabilidade da média.

Pelo Teorema Central do Limite tem-se que $\bar{X} \sim N(\mu_{\bar{X}} = \mu_X; \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}})$, portanto, se H_0 for verdade, e admitindo-se que os triatomíneos no ambiente com $25^\circ C$ possuem distribuição do tempo médio de vida na fase adulta: $\bar{X} \sim N(\mu_{\bar{X}} = 224,8; \sigma_{\bar{X}} = \frac{86}{\sqrt{22}})$, sob H_0 .

Pode-se utilizar $Z_{\bar{X}}$ ou \bar{x}_{obs} para a tomada de decisão.

Região de rejeição e aceitação da hipótese H_0 .



Cálculo do tempo médio de vida na amostra de 22 triatomíneos, $\bar{x}_{obs} = 176$.

Cálculo do tempo médio observado em número de desvios padrão:

$$Z_{\bar{x}_{obs}} = \frac{\bar{x}_{obs} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{176 - 224,8}{18,3} = -2,67$$

Confrontar o valor da estatística do teste com a região de rejeição e aceitação de H_0 .

Como Z_{obs} está à esquerda de $Z_{\text{crítico}}$ (região de rejeição), decide-se por rejeitar H_0 .

Decisão

Rejeita-se H_0 .

Conclusão

Foi encontrada diferença estatisticamente significativa entre os tempos de vida de triatomíneos submetidos a temperaturas de 25°C e 30°C a um nível de significância $\alpha = 0,05$. Temperatura maior promove um tempo de duração do triatomíneo, em fase adulta, menor.

Regra geral:

Rejeita-se H_0 se

$$Z_{obs} > Z_{crítico}$$

$$\text{para } H_a : \mu_{SDIG} > \mu_{Sadias}$$

$$Z_{obs} < -Z_{crítico}$$

$$\text{para } H_a : \mu_{SDIG} < \mu_{Sadias}$$

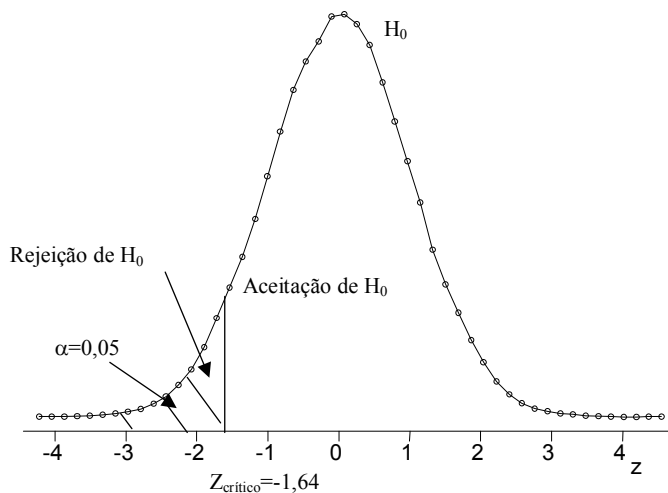
$$Z_{obs} > Z_{crítico} \text{ OU } Z_{obs} < -Z_{crítico}$$

$$\text{para } H_a : \mu_{SDIG} \neq \mu_{Sadias}$$

Se a estatística do teste for $\bar{x}_{observada}$

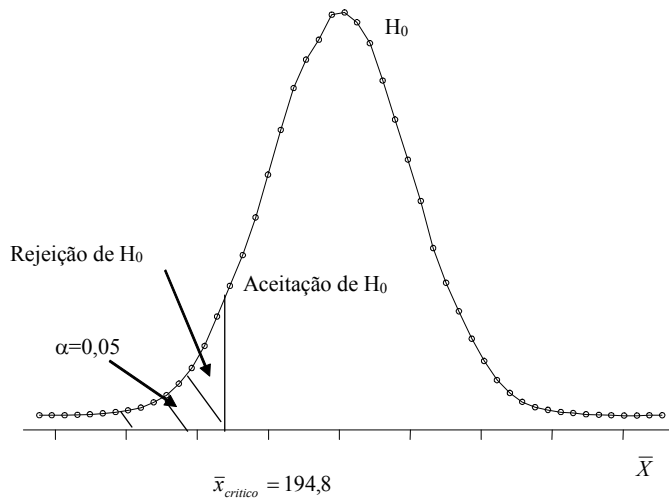
É possível realizar o teste comparando a média observada na amostra ($\bar{x}_{obs} = 176$) e o valor de tempo médio na fase adulta que deixa, no caso deste exemplo, uma área $\alpha = 0,05$ à sua esquerda. O valor de peso médio que limita esta área é denominado $\bar{x}_{crítico}$.

Sabe-se que para $\alpha = 0,05$, o valor de $Z_{crítico} = -1,64$.



Pode-se calcular o valor de $\bar{x}_{crítico}$ que limita a área de 5% utilizando a estatística $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma_X}{\sqrt{n}}}$

Substituindo-se os valores, $-1,64 = \frac{\bar{x} - 224,8}{18,3}$, então $\bar{x} = -1,64 \times 18,3 + 224,8 = 194,8$



Tomando-se a decisão

Comparando-se $\bar{x}_{obs} = 176$ e $\bar{x}_{critico} = 194,8$ pode-se ver que $\bar{x}_{obs} < \bar{x}_{critico}$ indicando que a hipótese H_0 deve ser rejeitada.

Regra geral para a tomada de decisão utilizando a estatística $\bar{x}_{observada}$:

Rejeita-se H_0 se

$$\bar{X}_{obs} > \bar{X}_{critico}$$

para $H_a : \mu_{30^\circ C} > \mu_{25^\circ C}$

$$\bar{X}_{obs} < -\bar{X}_{critico}$$

para $H_a : \mu_{30^\circ C} < \mu_{25^\circ C}$

$$\bar{X}_{obs} < -\bar{X}_{critico} \text{ ou } \bar{X}_{obs} > \bar{X}_{critico}$$

para $H_a : \mu_{30^\circ C} \neq \mu_{25^\circ C}$

Teste de hipóteses de uma média populacional (μ) (com variância conhecida)

Abordagem de Fisher

Tomando-se como exemplo os dados de duração das fases de evolução (dias) de *Triatoma sordida* publicados, em condições de laboratório por Juarez E. Silva EPC (1982) que descrevem que com temperaturas mais altas há um encurtamento do período como adulto, e se for considerado que a uma temperatura de 25°C, a duração na fase adulta segue distribuição normal com média populacional (μ) 224,8 dias e desvio padrão populacional (σ) 86,0 dias, pode-se realizar o experimento que consiste em expor os insetos a uma temperatura maior e quantificar o tempo de duração na fase adulta.

Para tanto foi realizado um estudo onde se observou o tempo médio na fase adulta de exemplares a uma temperatura de 30°C obtendo-se os tempos a seguir.

Deseja-se saber se existe efeito de temperatura na duração da fase adulta deste inseto

A variável de estudo X é tempo (dias) – variável resposta, desfecho ou outcome.

Tomando-se como referência a temperatura de 25°C, $X \sim N(\mu = 224,8; \sigma = 86,0)$.

Fisher recomenda que para a tomada de decisão se calcule o valor de p e a decisão é feita com base neste valor.

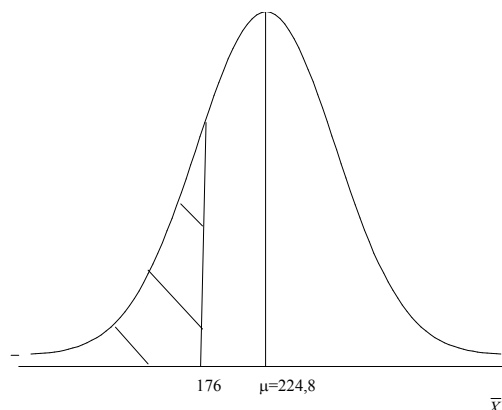
Valor de p é a probabilidade de observar a estatística do teste ou um valor mais extremo em uma curva especificada na proposição inicial.

Proposição inicial: O tempo médio de duração de triatomíneos na fase adulta é 224,8 dias.

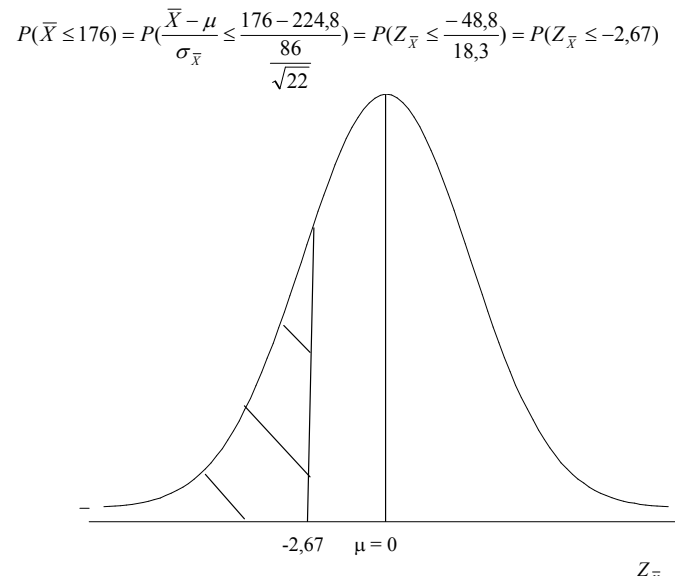
Distribuição de probabilidade:

Distribuição do tempo médio é uma distribuição normal com média $\mu = 224,8$ dias e desvio padrão

$$\frac{\sigma}{\sqrt{n}} = \frac{86}{\sqrt{22}} = 18,3$$



Cálculo da probabilidade de observar tempo médio na fase adulta igual ou menor que 176 na curva especificada na proposição inicial.



Pela distribuição Normal reduzida tem-se que $P(Z \leq -2,67) = 0,5 - 0,49621 = 0,00379$ ou 0,038%

Os resultados na temperatura de 25°C não são compatíveis com uma distribuição que tem tempo médio igual a 224,8 (temperatura de 30°C). A amostra vem de uma população que apresenta tempo médio na fase adulta menor que 224,8 ($p=0,004$).

Exemplo 64

Em 2006 foi realizado um estudo em 146 municípios brasileiros que utilizavam ações de controle para o *Aedes aegypti*. Se estes forem considerados como o universo de municípios, observou-se índice de infestação predial médio $\mu = 1,7\%$ e desvio padrão $\sigma = 0,45\%$. Em novo levantamento realizado em uma amostra de 10 destes municípios, em 2015, observou-se índice de infestação predial médio $\bar{x} = 1,9\%$. Realize um teste de hipóteses para investigar se a diferença observada ocorreu somente devido ao acaso ou se as médias populacionais provêm de populações diferentes. Utilize nível de significância de 5%. Tome a decisão também utilizando a abordagem de Fisher.

Exemplo 65

Sabe-se que o peso seco médio de *Anopheles darlingi* em laboratório é $\mu = 0,20mg$ com desvio padrão $\sigma = 0,05mg$. Em uma amostra de 50 exemplares que foram submetidos a restrição alimentar, obteve-se peso seco médio $\bar{x} = 0,28mg$. Seria razoável concluir que a verdadeira média de peso seco desta espécie sob restrição alimentar é a mesma que a média populacional da espécie que não sofreu restrição? Realize um teste de hipóteses segundo a abordagem de Neyman e Pearson para $\alpha = 0,05$ e a de Fisher para responder a pergunta. Calcule o tamanho da amostra necessária para responder tal questão fixando-se $\alpha = 0,05$; $\beta = 0,10$ e $d=0,3$.

Teste de uma média com variância populacional desconhecida

Supor o mesmo exemplo utilizado anteriormente onde se deseja comparar o tempo médio na fase adulta de triatomíneos criados em laboratório a uma temperatura de 30°C onde se conhecia o tempo médio populacional se a temperatura fosse 25°C. A diferença agora é que se desconhece o valor da variância populacional nesta temperatura. Então, tomando-se como referência a temperatura de 25°C, $X \sim N(\mu = 224,8; \sigma = ?)$.

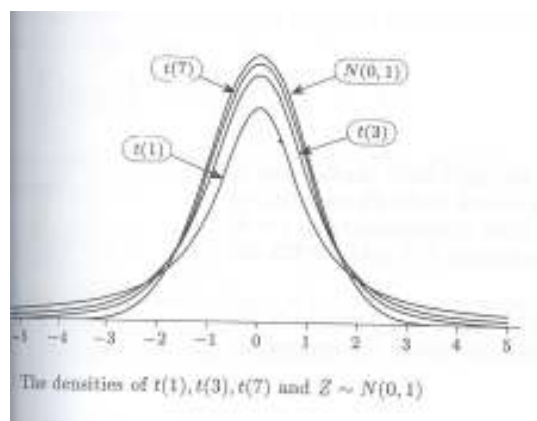
Nesta situação não é possível utilizar a curva normal pois esta depende dos parâmetros média e desvio padrão. Para resolver tal situação deve-se utilizar outra curva de probabilidade, a curva t de Student.

A família de distribuições t de Student

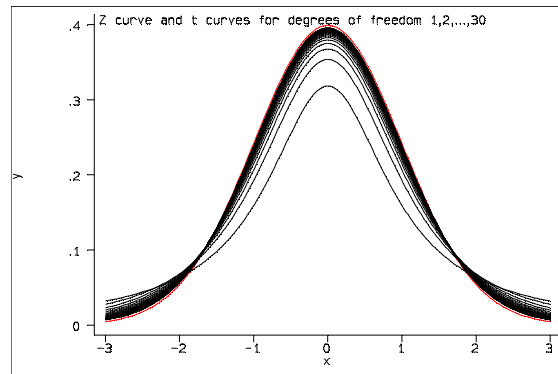
Student é o pseudônimo de William Sealy Gosset que, em 1908, propôs a distribuição t. Esta distribuição é muito parecida com a distribuição normal. A família de distribuições t é centrada no zero e possui formato em sino. A curva não é tão alta quanto a curva da distribuição normal e as caudas da distribuição t são mais altas que as da distribuição normal. O parâmetro que determina a altura e largura da distribuição t depende do tamanho da amostra (n) e é denominado graus de liberdade (gl),

denotado pela letra grega (ν) (lê-se ni). A notação da distribuição t é t_ν .

Curvas t para graus de liberdade (tamanhos de amostra) diferentes.



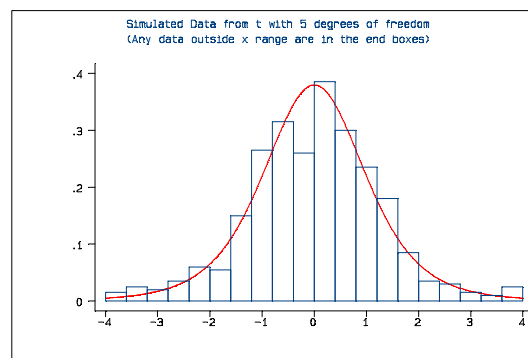
Quando o número de graus de liberdade da distribuição t aumenta, esta distribuição se aproxima de uma distribuição normal.



Esta família t não descreve o que acontece na natureza mas sim o que aconteceria se selecionássemos milhares de amostras aleatórias de uma população normal com média

$N(\mu = 0; \sigma = 1)$, e fosse calculado $t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$ para cada amostra.

Calculando-se o valor de t para 500 amostras de tamanho 6 de uma população com distribuição normal, obtém-se o gráfico a seguir:



Teste de uma média com variância populacional desconhecida

Supor a situação anterior, só que a variância (desvio padrão) populacional do tempo de estágio adulto quando submetido a uma temperatura de 25°C é desconhecida sendo conhecido somente o tempo médio populacional ($\mu_{25^\circ C} = 224,8$ dias).

Formulação das hipóteses

$$H_0 : \mu_{30^\circ C} = 224,8$$

$$H_a : \mu_{30^\circ C} < 224,8$$

Fixando-se o nível de significância $\alpha = 0,05$

Distribuição de probabilidade

Como as hipóteses envolvem a média populacional, é necessário utilizar a distribuição de probabilidade da média.

Pelo Teorema Central do Limite tem-se que $\bar{X} \sim N(\mu_{\bar{X}} = \mu_X; \sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}})$.

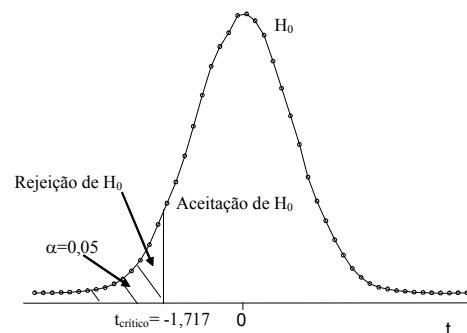
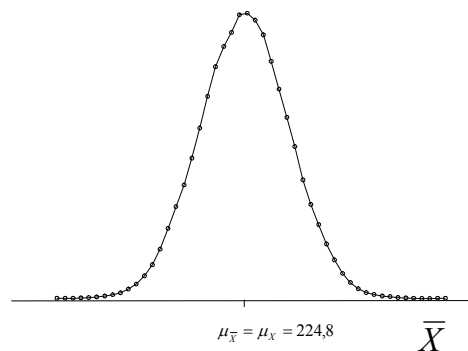
Admitindo-se que H_0 é verdade, resta um problema que é o fato de não se conhecer o valor da dispersão do tempo de estágio adulto em condições de temperatura igual a 25°C. Neste caso não é possível utilizar a estatística Z.

Utiliza-se, então, a estatística T onde $T = \frac{\bar{X} - \mu_{\bar{X}}}{S_{\bar{X}}} = \frac{\bar{X} - \mu_{\bar{X}}}{\frac{S_X}{\sqrt{n}}}$ sendo S_X o desvio padrão da população

de estudo, estimado com os dados da amostra de triatomas sordida.

T segue uma distribuição *t* de *Student*, com (n-1) graus de liberdade. Quando o tamanho da amostra é grande, a estatística T tende para uma distribuição normal com média 0 e desvio padrão 1 ($n \rightarrow \infty \Rightarrow T \sim N(0;1)$).

Região de rejeição e aceitação da hipótese H_0 .



Cálculo da estatística do teste

$$t_{\text{observado}} = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{176 - 224,8}{\frac{s}{\sqrt{22}}}$$

s é o desvio padrão dos dados amostrais, $s=47,55$

136	157	154	135	247
164	140	126	247	139
139	148	221	248	150
139	135	143	249	173
241	241			

$$t_{\text{observado}} = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{176 - 224,8}{\frac{47,55}{\sqrt{22}}} = -4,81$$

Como o valor de t calculado cai na área de rejeição de H_0 , decide-se por rejeitar H_0 .

Conclusão: foi encontrada diferença estatisticamente significativa entre o tempo médio no estágio adulto na temperatura de 25°C e 30°C para $\alpha = 0,05$. A temperatura maior provoca um tempo menor em fase adulta.

Teste de hipóteses para uma média populacional com variância desconhecida - Abordagem de Fisher

Supor a mesma situação anterior onde o tempo médio populacional de *Triatoma sordida* no estágio adulto para temperatura de 25°C é de 224,8 dias. Entretanto, o valor da variância populacional nesta temperatura é desconhecido. Resumindo-se, $X \sim N(\mu = 224,8; \sigma = ?)$. Deseja-se investigar o efeito da temperatura na duração deste estágio.

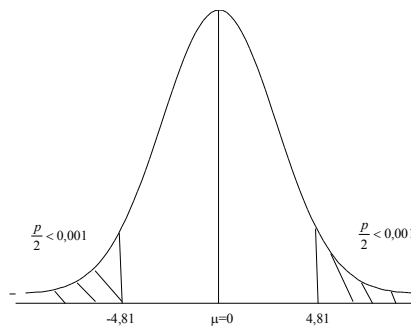
Proposição inicial: A espécie *Triatoma sordida* tem tempo médio de vida em estágio adulto igual a 224,8 dias.

Seleciona-se uma amostra de 22 espécimens e observa-se, em laboratório o tempo médio de vida em fase adulta quando expostos a temperatura de 30°C e obtem-se tempo médio \bar{x} igual a 176 e desvio padrão (s) igual a 47,55 dias.

Distribuição de probabilidade da estatística do teste:

$$t_{\text{observado}} = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}} = \frac{176 - 224,8}{\frac{47,55}{\sqrt{22}}} = -4,81$$

Estatística t (distribuição t de Student) com $n-1=22-1=21$ graus de liberdade.



Pela distribuição *t de Student* com 21 graus de liberdade, tem-se $P(t_{obs} \leq -4,81 \text{ ou } t_{obs} \geq 4,81) < 0,1\%$

Os resultados não são compatíveis com uma distribuição que tem tempo médio igual a 224,8 dias. Pode-se dizer que *Triatoma sordida* em temperatura mais alta tem seu tempo em estágio adulto diminuído ($p < 0,001$).

Exemplo 66

A literatura fornece informações a respeito do comprimento da asa de *Aedes scapularis* capturados no município de São Paulo como sendo $\mu = 2,5mm$. Em estudo realizado no município de Pariquearaçu, obteve-se os seguintes valores para esta característica

2,8	2,5	2,7	2,7	2,8	2,7	2,2
2,5	2,3	2,4	2,5	2,9	2,6	2,4
2,7	2,3	2,8	2,3	2,6	2,9	2,5

Realize um teste de hipóteses para comparar o comprimento da asa de *Aedes scapularis* coletados em Pariquearaçu com os dados de literatura do município de São Paulo. Utilize a abordagem clássica de Neyman e Pearson com nível de significância de 5%.

Tome a decisão também pela abordagem de Fisher. Calcule o valor de p e discuta os resultados.

Exemplo 67

Ninfas de *Triatoma klugi* apresentam duração média em estágio II igual a $\mu = 20,9 \text{ dias}$. Ao estudarem 88 ninfas em estágio II, pesquisadores observaram tempo médio (\bar{x}) neste estágio igual a 23,9 dias e desvio padrão (s) igual a 6,27 dias. Realize um teste de hipóteses para investigar se ninfas em estágio II e III apresentam tempo médio semelhante. Utilize a abordagem clássica de Neyman e Pearson com nível de significância de 5%.

Tome a decisão também pela abordagem de Fisher. Calcule o valor de p e discuta os resultados.

Exemplo 68

Segundo dados de literatura, é esperado que o volume médio de água em bromélias favoreça a presença de *Aedes albopictus*. Em bromélias sem presença deste mosquito de Ilhabela, litoral norte do estado de São Paulo, é descrito na literatura, quantidade média de água igual a $\mu = 238\text{ml}$. Em amostra de 30 bromélias com presença de *Aedes albopictus*, observou-se quantidade média de água (\bar{x}) igual a 275 ml e desvio padrão (s) igual a 80 ml. Realize um teste de hipóteses para investigar se a diferença na quantidade média de água em bromélias com e sem a presença deste mosquito ocorre somente devido ao acaso. Utilize a abordagem clássica de Neyman e Pearson com nível de significância de 5%.

Tome a decisão também pela abordagem de Fisher. Calcule o valor de p e discuta os resultados.

Exemplo 69

A abundância média de *Aedes albopictus* em Ilhabela na mata é de $\mu = 2,28$ imaturos por planta. Em levantamento realizado em 20 domicílios da área periurbana, observou-se abundância média (\bar{x}) de 2,12 imaturos por planta e desvio padrão (s) igual a 0,78 imaturos por planta. Realize um teste de hipóteses para investigar se existe diferença na abundância média na área periurbana e urbana. Utilize a abordagem clássica de Neyman e Pearson com nível de significância de 5%.

Tome a decisão também pela abordagem de Fisher. Calcule o valor de p e discuta os resultados.

Teste de hipóteses de associação pelo Qui-quadrado de Pearson (χ^2)

Qui-quadrado de Pearson – indica se há ou não associação. Não mede força de associação.

Duas variáveis qualitativas

X - curso universitário e

Y – sexo do aluno

Questão: sexo do indivíduo influi na escolha do curso?

Situação 1

Curso	Masculino n	Feminino n	Total n
Economia	24	36	60
Administração	16	24	40
Total	40	60	100

Curso	Masculino		Feminino		Total	
	n	proporção	n	proporção	n	proporção
Economia	24	0,6	36	0,6	60	0,6
Administração	16	0,4	24	0,4	40	0,4
Total	40	1	60	1	100	1

As proporções de escolha dos cursos não diferem segundo sexo do estudante.

Situação 2

Curso	Masculino n	Feminino n	Total n
Física	100 (a)	20 (b)	120
Ciências Sociais	40 (c)	40 (d)	80
Total	140	60	200

Curso	Masculino		Feminino		Total	
	n	proporção	n	proporção	n	proporção
Física	100	0,7	20	0,3	120	0,6
Ciências Sociais	40	0,3	40	0,7	80	0,4
Total	140	1	60	1	200	1

A distribuição de alunos em cada curso segundo sexo não é a mesma. Sexo e curso podem estar associados.

Se a variável sexo não fosse associada à escolha do curso, quantos indivíduos esperaríamos em Física, entre os homens?

Curso	Sexo	Número esperado
Física	Masculino (a)	$140 \times 0,6 = 140 \times \frac{120}{200} = 84$
Física	Feminino (b)	$60 \times 0,6 = 60 \times \frac{120}{200} = 36$
Ciências Sociais	Masculino (c)	$140 \times 0,4 = 140 \times \frac{80}{200} = 56$
Ciências Sociais	Feminino (d)	$60 \times 0,4 = 60 \times \frac{80}{200} = 24$

Tabela de **freqüências esperadas**, sob a condição de independência

Curso	Masculino	Feminino	Total
	n	n	n
Física	84	36	120
Ciências Sociais	56	24	80
Total	140	60	200

Valores observados O	Valores esperados E	(O-E)	(O-E) ²	$\frac{(O-E)^2}{E}$
100	84	16	256	3,048
40	56	-16	256	4,571
20	36	-16	256	7,11
40	24	16	256	10,667

Qui-quadrado=25,397

O Qui-quadrado é obtido somando-se a diferença ao quadrado entre as freqüências observadas e as esperadas, dividido pelas freqüências esperadas

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

Se o Qui-quadrado for igual a zero, então não existe associação entre as variáveis. O teste de associação pelo qui quadrado, portanto, investiga se o valor da estatística do teste é suficientemente distante do zero para se afirmar que existe associação entre as variáveis.

Exemplo

Um estudo investiga se a classificação do caso de dengue quanto ao local onde o indivíduo contraiu a doença está associado ao sexo.

Distribuição de casos de dengue segundo tipo de caso definido pelo local onde a pessoa contraiu a doença e sexo do paciente. Município de São Paulo, 2003.

Casos	Masculino		Feminino		Total
	n		n		n
Autóctones	380	52,6	407	48,3	787
Importados	342	47,4	436	51,7	778
Total	722	100	843	100	1565

Cálculo do qui-quadrado de Pearson

Valores observados O	Valores esperados E	(O-E)	(O-E) ²	$\frac{(O-E)^2}{E}$
380	363,08	16,92	286,29	0,789
342	358,92	-16,92	286,29	0,798
407	423,92	-16,92	286,29	0,675
436	419,08	16,92	286,29	0,683
Qui-quadrado=2,945				

O qui-quadrado é obtido somando-se razões dadas pelos quadrados das diferenças entre frequências observadas e as esperadas, divididos pelas frequências esperadas.

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

Se as variáveis forem independentes, então é equivalente a dizer que não existe associação, e neste caso, o valor do qui-quadrado será zero. O qui-quadrado não mede força de associação e não é suficiente para estabelecer relação de causa e efeito. Entretanto, a diferença em relação ao zero pode ter ocorrido somente devido ao acaso. Para a tomada de decisão sobre a existência de associação com significado estatístico, realiza-se o teste de hipóteses.

Abordagem de Neyman e Pearson

Estabelecimento das hipóteses:

$$\begin{cases} H_0: \text{Não existe associação} \\ H_a: \text{Existe associação} \end{cases}$$

Fixando-se a probabilidade de erro tipo I:

Nível de significância (α) = 0,05

Estatística do teste:

$$\text{Qui-quadrado} = \sum \frac{(O - E)^2}{E} \sim \chi^2_{(r-1)(c-1)}$$

onde r e c representam o número de linhas e de colunas, respectivamente.

Correção de continuidade:

$$\text{Qui-quadrado}_{\text{correcao de Yates}} = \sum \frac{(|O - E| - 0,5)^2}{E} \sim \chi^2_{(r-1)(c-1)}$$

Limitações:

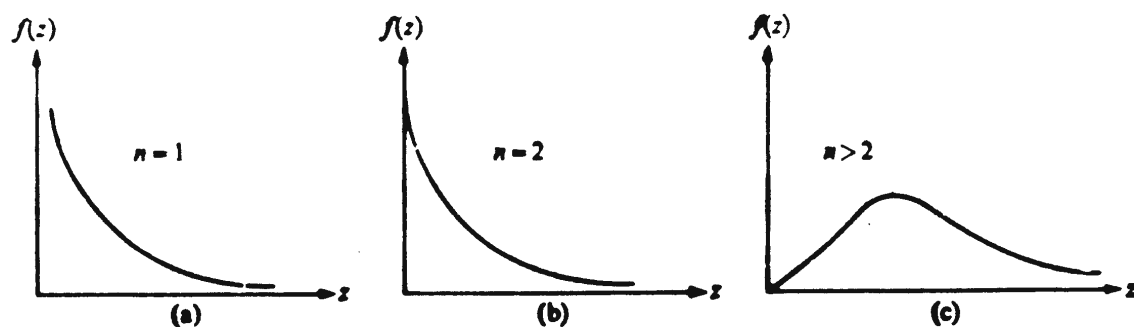
Para $n < 20$, utilizar o teste exato de Fisher

Para $20 \leq n \leq 40$, utilizar o qui-quadrado somente se os valores esperados forem maiores ou iguais a 5

Distribuição qui-quadrado ($\chi^2_{(n-1)}$) com (n-1) graus de liberdade

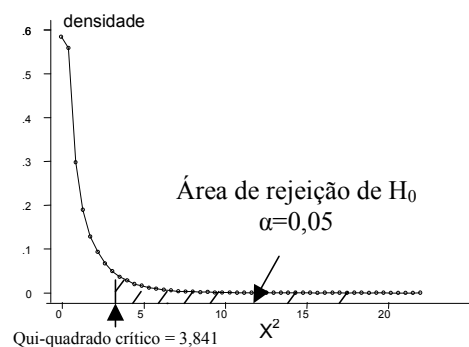
Seja uma população com distribuição normal $N(\mu, \sigma)$. Se desta população se obtiver um número infinito de amostras de tamanho n , calculando-se as quantidades \bar{X} e S^2 em cada amostra, a variável aleatória $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{(n-1)}$, onde $\chi^2_{(n-1)}$ se lê "qui-quadrado com n-1 graus de liberdade" Berquó (1981).

A distribuição qui-quadrado é assimétrica e se torna menos assimétrica a medida que os graus de liberdade aumentam. Os valores da distribuição são sempre positivos (maior ou igual a zero). Existe uma família de distribuições qui-quadrado, dependendo do número de graus de liberdade. Para grandes amostras, a distribuição qui-quadrado tende para uma distribuição normal.



Definição das áreas de aceitação e rejeição de H_0

Para a tomada de decisão, utiliza-se a regra: rejeita-se H_0 se o valor calculado do qui-quadrado for maior do que o valor crítico para um nível de significância pré definido.



Qui-quadrado calculado = 2,945

Como este valor cai na área de aceitação, decide-se por não rejeitar H_0 .

Conclui-se que as variáveis sexo do paciente e local onde contraiu a doença não estão estatisticamente associados para nível de significância de 5%.

Exemplo

Distribuição de flebotomíneos coletados em armadilhas com luz colocadas a 3 pés e 35 pés de altura em relação ao solo em uma área do leste do Panamá, 1972.

Pés*	Macho		Fêmea		Total	
	n	%	n	%	n	%
3 (0,914cm)	173	53,6	150	46,4	323	100
35 (1066,8 cm)	125	63,1	73	36,9	198	100
Total	298	57,2	223	42,8	521	100

*1 pé=30,48 cm

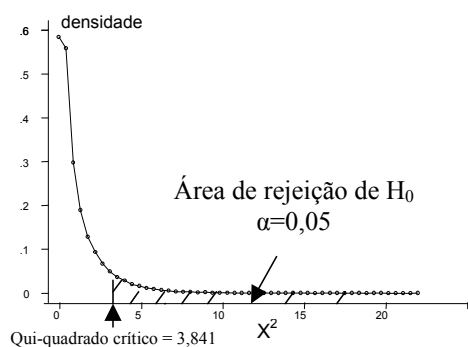
Fonte: Hand DJ et al. A handbook of small data sets. Chapman & Hall, 1994.

Estabelecimento das hipóteses:

$$\begin{cases} H_0: \text{Não existe associação} \\ H_a: \text{Existe associação} \end{cases}$$

Fixando-se a probabilidade de erro tipo I:

Nível de significância (α) = 0,05



Cálculo do qui-quadrado de Pearson

Valores observados O	Valores esperados E	(O-E)	(O-E)²	$\frac{(O-E)^2}{E}$	$\frac{(O-E -0,5)^2}{E}$
173	184,7	-11,7	136,89	0,741	0,679
125	113,3	11,7	136,89	1,208	1,107
150	138,3	11,7	136,89	0,990	0,907
73	84,7	-11,7	136,89	1,616	1,481
				$\chi^2 = 4,555$	$\chi^2_{\text{corrigido}} = 4,174$

Coeficiente de associação de Yule – permite investigar a força (magnitude) da associação

Para uma tabela 2x2

X	Y ₁	Y ₂	Total
	n	n	n
X ₁	(a)	(b)	m ₁
X ₂	(c)	(d)	m ₂
Total	n ₁	n ₂	n

$$Y = \frac{a.d - b.c}{a.d + b.c}, \text{ onde: } -1 \leq Y \leq +1$$

$$Y = \frac{173 \times 73 - 125 \times 150}{173 \times 73 + 125 \times 150} = -0,195$$

Como o valor do qui-quadrado corrigido caiu na área de rejeição, decide-se por rejeitar H₀. As variáveis altura da armadilha e o sexo do flebotômico estão associados. Fêmeas estão presentes em menor proporção em alturas maiores do que os machos para nível de significância de 5%. Pelo valor do Coeficiente de Yule pode-se afirmar que esta associação é fraca.

Exercício 70

Distribuição de culturas para *Leptospira* em amostras em animais e ambiente. Egito, 2015

Tipo de fonte	Cultura positiva		Cultura negativa		Total	
	n	%	n	%	n	%
Ratos	17		253		270	
Cachorros	19		149		168	
Vacas	7		618		625	
Recursos com água	10		35		45	
Total	53		1055		1108	

Fonte: samir A et al. Ver Sociedade Brasileira de medicina Tropical 48(3):272-277,2015

Investigue a existência de associação entre as variáveis. Utilize a abordagem clássica de Neyman e Pearson com nível de significância de 5%. Discuta os resultados.

Abordagem de Fisher

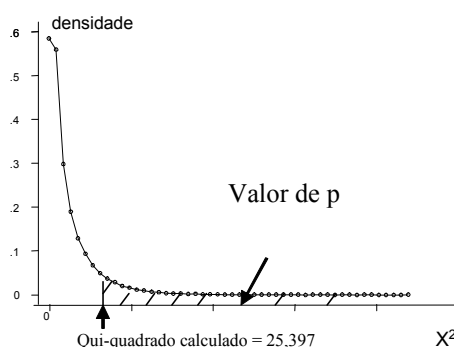
Situação 2

Curso	Masculino	Feminino	Total
	n	n	n
Física	100 (a)	20 (b)	120
Ciências Sociais	40 (c)	40 (d)	80
Total	140	60	200

Proposição inicial: Não existe associação

Qui quadrado calculado = 25,397

Para gl=1, p<0,001



Os dados mostram evidência contrária à proposição inicial, a escolha do curso depende do sexo do aluno ($p < 0,001$). Estudantes do sexo masculino escolhem mais Física do que os do sexofeminino.

Exercício 71

Abordagem de Fisher para a tomada de decisão sobre a existência de associação entre as variáveis

Distribuição de casos de dengue segundo tipo de caso definido pelo local onde a pessoa contraiu a doença e sexo do paciente. Município de São Paulo, 2003.

Casos	Masculino		Feminino		Total	
	n		n		n	
Autóctones	380	52,6	407	48,3	787	50,3
Importados	342	47,4	436	51,7	778	49,7
Total	722	100	843	100	1565	100

Qui quadrado calculado = 2,945

Proposição inicial:

Valor de p:

Conclusão:

Exercício 72

Abordagem de Fisher para a tomada de decisão sobre a existência de associação entre as variáveis

Distribuição de flebotomíneos coletados em armadilhas com luz colocadas a 3 pes e 35 pés de altura em relação ao solo em uma área do leste do Panamá, 1972.

Pés*	Macho		Fêmea		Total	
	n	%	n	%	n	%
3 (0,914cm)	173	53,6	150	46,4	323	100
35 (1066,8 cm)	125	63,1	73	36,9	198	100
Total	298	57,2	223	42,8	521	100

*1 pé=30,48 cm

Fonte: Hand DJ et al. A handbook of small data sets. Chapman & Hall, 1994.

Qui quadrado bruto = 4,555

Proposição inicial:

Valor de p:

Conclusão:

Qui quadrado corrigido = 4,174

Proposição inicial:

Valor de p:

Conclusão:

Exercício 73

Abordagem de Fisher para a tomada de decisão sobre a existência de associação entre as variáveis

Distribuição de culturas para *Leptospira* em amostras em animais e ambiente. Egito, 2015

Tipo de fonte	Cultura positiva		Cultura negativa		Total	
	n	%	n	%	n	%
Ratos	17		253		270	
Cachorros	19		149		168	
Vacas	7		618		625	
Recursos com água	10		35		45	
Total	53		1055		1108	

Fonte: samir A et al. Ver Sociedade Brasileira de medicina Tropical 48(3):272-277,2015

Qui quadrado =

Proposição inicial:
Valor de p:
Conclusão:

Estimação de parâmetros populacionais

Estimação por ponto

X é uma característica que na população possui distribuição normal com média μ e variância σ^2 (desvio padrão σ).

Seja $X_1, X_2, X_3, \dots, X_n$ uma amostra aleatória de tamanho **n** extraída desta população.

Os parâmetros μ e σ^2 podem ser estimados com base na amostra.

Se o estimador for um único valor, a estimação é chamada de estimação por ponto.
Se o estimador for um conjunto de valores, a estimação é chamada de estimação por intervalo.

Estimação por ponto

Média aritmética

Populacional Parâmetro μ estimador : $\bar{X} = \frac{\sum_{i=1}^N X_i}{N}$

Variância

Populacional Parâmetro σ^2 estimador : $S_{(N)}^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N}$ ou $S_{(N-1)}^2 = \frac{\sum_{i=1}^N (X_i - \bar{X})^2}{N-1}$

Atenção: Antes dos dados serem coletados, **os estimadores são variáveis aleatórias**.

Estimação por intervalo

Intervalo de confiança: É um conjunto de valores calculados com base na amostra. Pressupõe-se que cubra o parâmetro de interesse com um certo grau (nível) de confiança.

O grau de confiança tem origem na probabilidade associada ao processo de construção do intervalo antes de se obter o resultado amostral.

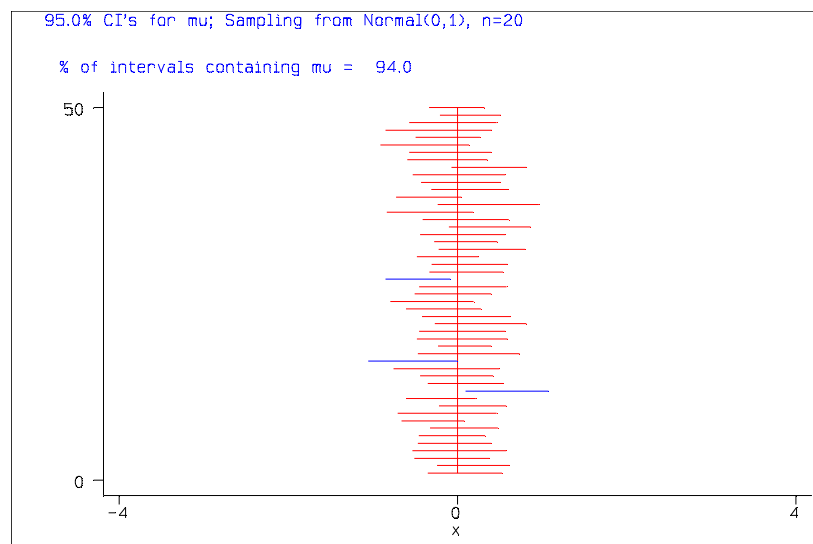
O grau de confiança mais comumente utilizado é o de 95%.

Seria impossível construir um intervalo de 100% de confiança a menos que se medisse toda a população.

Na maioria das aplicações não sabemos se um intervalo de confiança específico cobre o verdadeiro valor. Só podemos aplicar o conceito frequentista de probabilidade e dizer que se realizarmos a amostragem infinitas vezes e construirmos intervalos de confiança de 95%, em 95% das vezes os intervalos de confiança estarão corretos (cobrirão o parâmetro) e 5% das vezes estarão errados.

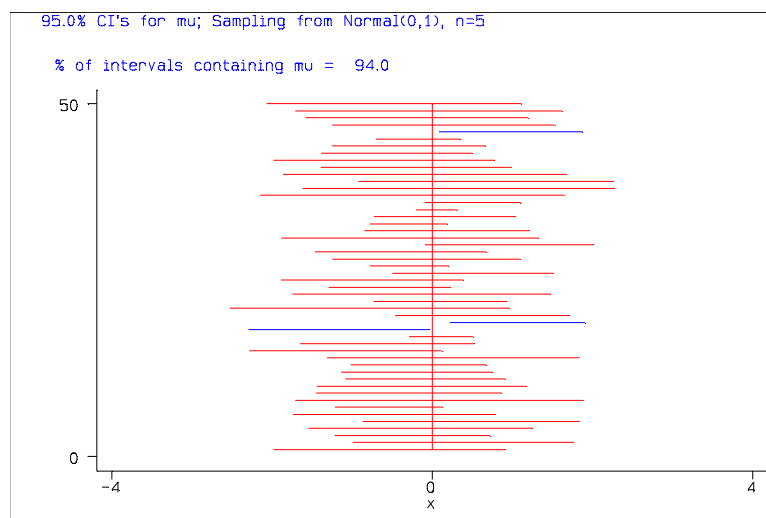
Exemplos de intervalo de confiança:

Representação gráfica



A linha vertical representa o parâmetro populacional. O gráfico foi gerado via programa de computador. São apresentados 50 intervalos de confiança para amostras de tamanho $n=20$. As linhas horizontais representam os intervalos de confiança. Se o intervalo de confiança não contiver o parâmetro, a linha horizontal não cruzará a linha vertical. A linha vertical é o parâmetro. No exemplo, 3 intervalos não cobrem ("capturam") o parâmetro.

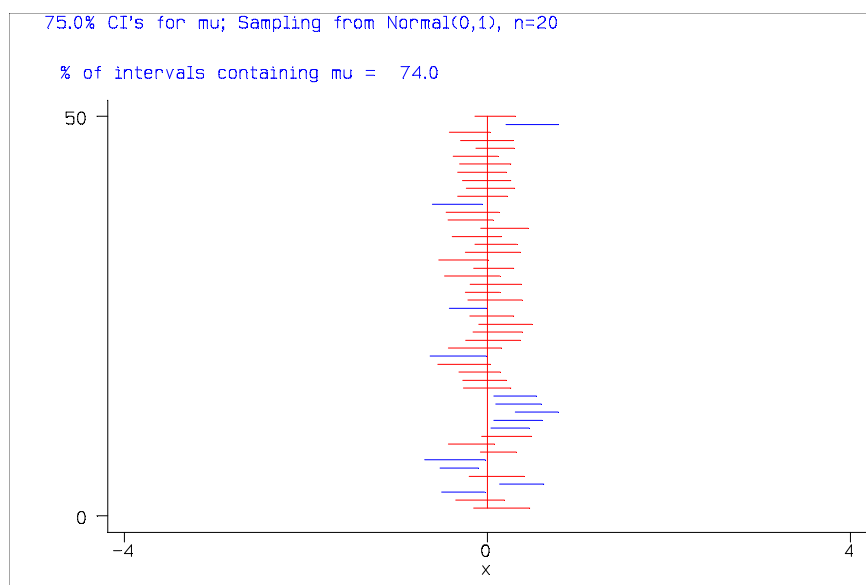
Apresentação gráfica do efeito do tamanho da amostra:



Para amostras menores ($n=5$), as larguras dos intervalos são maiores a proporção de intervalos que "capturam" o parâmetro é parecida com a anterior (para $n=20$). Portanto, o tamanho da amostra não interfere na proporção de "captura" do parâmetro mas sim na precisão do estimador.

Efeito do grau de confiança

Para $n=20$ e $\alpha=0,25$, obtém-se intervalos com os apresentados a seguir:



Os intervalos são mais estreitos do que para $n=20$ e $\alpha=0,05$. Uma porcentagem bem maior não contém o parâmetro. Isto é o que 75% de confiança significa. Do total de todas as possíveis amostras, 75% delas resultará em intervalos de confiança que contêm o verdadeiro valor do parâmetro.

Interpretando Intervalos de Confiança (IC)

Um intervalo de confiança para um parâmetro é um intervalo de valores no qual pode-se depositar uma confiança que o intervalo cobre (contém) o valor do parâmetro.

Intervalo de confiança para a proporção de larvas *Anopheles albitarsis* s.l. (π): Em estudo realizado para estimar a proporção populacional de larvas *Anopheles albitarsis* s.l. num determinado lago coletou-se 1000 larvas de mosquitos. Destes, 250 eram da espécie de interesse. $\hat{p} = 0,25$ ou 25% e o IC(95%) será (22,3% - 27,7%). O valor 25% é a proporção estimada no ponto e o intervalo (22,3% - 27,7%) estima o parâmetro com grau de confiança de 95%. Deposita-se neste intervalo uma confiança de 95% que ele cobre (contém) o valor do parâmetro.

Intervalo de confiança para o peso seco médio populacional *Culex quinquefasciatus* (μ):

Ao se estudar o peso seco de *Culex quinquefasciatus* de uma amostra de 300 exemplares observou-se peso médio (\bar{x}) igual a 0,45 mg e desvio padrão populacional (σ) igual a 0,11 mg. Com base na amostra obtém-se o intervalo (0,438 ; 0,462 mg) que é um intervalo de 95% de confiança para a média (μ) da população de mosquitos desta espécie. Então podemos estar 95% confiantes que o conjunto de valores cobre (contém) o verdadeiro peso seco médio populacional.

Pode-se também pensar no IC a partir da seleção de milhares de amostras de uma população. Para cada amostra calcula-se um intervalo de confiança com grau de confiança $100(1-\alpha)\%$, para um parâmetro da população. A porcentagem de intervalos que contém o verdadeiro valor do parâmetro é $100(1-\alpha)$. Para $\alpha = 0,05$, o grau de confiança será igual a $100(1-0,05)\% = 100(0,95)\% = 95\%$.

Na prática, tomamos somente uma amostra e obtemos somente um intervalo. Mas sabemos que $100(1-\alpha)\%$ de todas as amostras tem um intervalo de confiança contendo o verdadeiro valor do parâmetro, portanto depositamos uma confiança $100(1-\alpha)\%$ que o particular intervalo contém o verdadeiro valor do parâmetro.

Amplitude do intervalo

Para um grau de confiança especificado (por exemplo, 95%), desejamos o intervalo tão pequeno quanto possível.

Ex: o intervalo de confiança de 95% para o peso seco médio (mg) de mosquitos *Culex quinquefasciatus* de (0,275 ; 0,894 mg) traz pouca informação prática porque a amplitude é muito grande apesar de ser um dos possíveis intervalos que captura a verdadeira média com grau de

confiança de 95%. Na prática, deseja-se um intervalo com amplitude pequena pois este estima um único valor. É o tamanho da amostra que determina a amplitude do intervalo. Quanto maior a amostra, menor será o intervalo.

Fórmulas para construção dos intervalos de confiança:

As fórmulas dos intervalos de confiança são derivadas da distribuição amostral da estatística;

Construção do intervalo de confiança para a média populacional μ ;

Pressuposição: A amostra deve ser obtida de forma aleatória;

É necessário utilizar as propriedades do teorema central do limite :

$$X \sim N(\mu, \sigma); \bar{X} \sim N(\mu, \frac{\sigma}{\sqrt{n}})$$

Padronizando-se a média \bar{X} , obtém-se $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$, que permite calcular

$$P(-z \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z) = 1 - \alpha.$$

$$\text{Para } \alpha = 5\%, P(-1,96 \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq +1,96) = 0,95$$

$$P(-1,96 \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq +1,96 \frac{\sigma}{\sqrt{n}}) = 0,95$$

$$P(-\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} \leq -\mu \leq -\bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}) = 0,95$$

Multiplicando tudo por -1

$$P(\bar{X} + 1,96 \frac{\sigma}{\sqrt{n}} \geq \mu \geq \bar{X} - 1,96 \frac{\sigma}{\sqrt{n}}) = 0,95$$

Reescrevendo a equação tem-se

$$P(\bar{X} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + 1,96 \frac{\sigma}{\sqrt{n}}) = 0,95$$

Obtém-se um intervalo aleatório **centrado na média amostral** o qual possui 95% de probabilidade de conter a verdadeira média populacional.

O parâmetro será estimado por um conjunto de valores provenientes de uma amostra. Quando isto é feito, a média é estimada por um determinado valor ($\hat{X} = \bar{x}$), e o intervalo $\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}}$ deixa de ser uma variável aleatória.

Este intervalo cobre (contém) ou não cobre (não contém) a verdadeira média (parâmetro). Diz-se então que a confiança que se deposita neste intervalo é de 95% porque antes de coletar a amostra de tamanho n , existia, associada a ele, uma probabilidade de 95% de que contivesse a média populacional. Por isso chama-se intervalo de confiança para a média populacional.

$$IC(95\%) : (\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}})$$

Intervalo de confiança para a média populacional com variância populacional conhecida

Pressuposição: A amostra deve ser obtida de forma aleatória.

Estatística: média populacional - μ .

$$IC(\mu) = \bar{x} - z_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}}; \bar{x} + z_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}}$$

Exemplo:

Ao se estudar o peso seco de *Culex quinquefasciatus* de uma amostra de 300 exemplares observou-se peso médio (\bar{x}) igual a 0,45 mg e desvio padrão populacional (σ) igual a 0,11 mg. Com base na amostra obtenha o intervalo de 95% de confiança para o peso médio (μ) da população.

$$IC(\mu) : (\bar{x} - z_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}}; \bar{x} + z_{\alpha/2} \cdot \frac{\sigma_x}{\sqrt{n}})$$

$$IC(\mu) : (0,45 - 1,96 \cdot \frac{0,11}{\sqrt{300}}; 0,45 + 1,96 \cdot \frac{0,11}{\sqrt{300}})$$

$$IC(95\%) : (0,45 - 0,012; 0,45 + 0,012)$$

$$IC(95\%) : (0,438; 0,462 \text{ mg})$$

Exercício 75

Estime o comprimento da asa de *Aedes scapularis* capturados no município de Pariqueiraçu com grau de confiança de 95%. Sabendo-se que o desvio padrão populacional (σ) é de 1,24 mm.

2,8	2,5	2,7	2,7	2,8	2,7	2,2
2,5	2,3	2,4	2,5	2,9	2,6	2,4
2,7	2,3	2,8	2,3	2,6	2,9	2,5

Exercício 76

Com base na amostra de 88 ninfas de *Triatoma klugi* observou-se tempo médio (\bar{x}) no estágio II igual a 23,9 dias. Sabendo-se que o desvio padrão populacional (σ) do tempo médio neste estágio é de 6,4 dias, estime o tempo médio no estágio, da população, com grau de confiança de 95%.

Exercício 77

Em estudo realizado em Ilhabela observou-se que em 20 bromélias com presença de *Aedes albopictus*, o volume médio de água (\bar{x}) foi igual a 275 ml. Considerando-se que o desvio padrão populacional (σ) é igual a 80 ml, estime o volume médio de água da população de bromélias com presença deste mosquito, com grau de confiança de 95%.

Exercício 78

Estime a abundância média de *Aedes albopictus* em área urbana de Ilhabela sabendo-se que em amostra de 20 domicílios, observou-se abundância média (\bar{x}) de 2,12 imaturos por planta. Considere o desvio padrão populacional (σ) como sendo igual a 0,78 imaturos por planta. Utilize grau de confiança de 90%.

Intervalo de confiança para a média populacional com variância populacional desconhecida

$$IC(\mu): \bar{x} - t_{n-1, \alpha/2} \cdot \frac{S_x}{\sqrt{n}}; \bar{x} + t_{n-1, \alpha/2} \cdot \frac{S_x}{\sqrt{n}}$$

Exemplo:

Ao se estudar o peso seco de *Culex quinquefasciatus* de uma amostra de 30 exemplares observou-se peso médio (\bar{x}) igual a 0,45 mg. O desvio padrão populacional (σ) era desconhecido e foi estimado na amostra obtendo-se o valor 0,19 mg. Com base na amostra obtenha o intervalo de 95% de confiança para o peso médio (μ) da população.

$$IC(\mu): (\bar{x} - t_{n-1; \alpha/2} \cdot \frac{S_x}{\sqrt{n}}; \bar{x} + t_{n-1; \alpha/2} \cdot \frac{S_x}{\sqrt{n}})$$

$$IC(\mu): (0,45 - 2,045 \cdot \frac{0,19}{\sqrt{30}}; 0,45 + 2,045 \cdot \frac{0,19}{\sqrt{30}})$$

Gl=n-1 = 29

IC(95%): (0,45 – 0,071; 0,45 + 0,071)

IC(95%): (0,479; 0,521 mg)

Exercício 79

Estime o comprimento da asa de *Aedes scapularis* capturados no município de Pariquearaçu com grau de confiança de 95%.

2,8	2,5	2,7	2,7	2,8	2,7	2,2
2,5	2,3	2,4	2,5	2,9	2,6	2,4
2,7	2,3	2,8	2,3	2,6	2,9	2,5

Exercício 80

Com base na amostra de 88 ninfas de *Triatoma klugi* observou-se tempo médio (\bar{x}) no estágio II igual a 23,9 dias. Sabendo-se que o desvio padrão amostral (s) do tempo médio neste estágio é de 5,7 dias, estime o tempo médio no estágio, da população, com grau de confiança de 95%.

Exercício 81

Em estudo realizado em Ilhabela observou-se que em 20 bromélias com presença de *Aedes albopictus*, o volume médio de água (\bar{x}) foi igual a 275 ml. Considerando-se que o desvio padrão amostral (s) é igual a 75 ml, estime o volume médio de água da população de bromélias com presença deste mosquito, com grau de confiança de 95%.

Exercício 82

Estime a abundância média de *Aedes albopictus* em área urbana de Ilhabela sabendo-se que em amostra de 20 domicílios, observou-se abundância média (\bar{x}) de 2,12 imaturos por planta. Considere o desvio padrão amostral (s) como sendo igual a 0,69 imaturos por planta. Utilize grau de confiança de 90%.

Intervalo de confiança aproximado pela Normal para a proporção populacional (π)

Pressuposições:

- 1- np e $nq \geq 5$
- 2- a amostra deve ser obtida de forma aleatória

Seja X uma variável aleatória que segue uma distribuição binomial, $X \sim B(n,p)$.

A proporção de sucessos populacional π é desconhecido. Seu estimador por ponto é $p = \frac{X}{n}$

Para n grande, $p \sim N(\mu = p, \sigma = \sqrt{\frac{pq}{n}})$.

Com intervalo de confiança para π dado por

$IC(\pi): \hat{p} - z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}; \hat{p} + z_{\alpha/2} \sqrt{\frac{\hat{p}\hat{q}}{n}}$, com \hat{p} e \hat{q} estimados na amostra onde $\hat{p} = \frac{x}{n}$ e $\hat{q} = 1 - \hat{p}$ e x é o número de sucessos observado na amostra de tamanho n .

Exemplo:

Intervalo de confiança para a proporção de larvas *Anopheles albitarsis* s.l. (π): Em estudo realizado para estimar a proporção populacional de larvas *Anopheles albitarsis* s.l. num determinado lago coletou-se 1000 larvas de mosquitos. Destes, 250 eram da espécie de interesse. $\hat{p} = 0,25$ ou 25% . Estime a proporção populacional com grau de confiança de 95%.

$$\hat{p} = \frac{250}{1000} = 0,25; \hat{q} = 1 - \hat{p} = 0,75$$

$$IC(95\%, \pi): 0,25 - 1,96 \sqrt{\frac{0,25 \times 0,75}{1000}}; 0,25 + 1,96 \sqrt{\frac{0,25 \times 0,75}{1000}}$$

$$IC(95\%, \pi): (0,223; 0,277) \text{ ou } (22,3\%; 27,7\%)$$

Resumo: Intervalo de Confiança

Média populacional: μ

Com variância conhecida σ^2 : $\bar{x} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{x} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

Com variância σ^2 desconhecida: $\bar{x} - t_{\alpha/2, v} \frac{s}{\sqrt{n}}; \bar{x} + t_{\alpha/2, v} \frac{s}{\sqrt{n}}; v = n - 1$

Proporção populacional (p)

Intervalo aproximado: $n\hat{p}, n\hat{p}(1 - \hat{p}) \geq 5$

$$\hat{p} - Z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}; \hat{p} + Z_{\alpha/2} \sqrt{\frac{\hat{p}(1 - \hat{p})}{n}}$$

onde \hat{p} é a proporção de sucessos na amostra.

Exercício 83

Deseja-se estimar a proporção populacional de fêmeas nulíparas de *Anopheles cruzii* com *habitat* no Parque estadual de Palmito, município de Paranaguá, Paraná. Considere que em amostra de 208 fêmeas, 120 eram nulíparas. Utilize grau de confiança de 95%.

Exercício 84

Deseja-se estimar a proporção populacional de pupas fêmeas de *Aedes aegypti* com *habitat* em pneus. Considere que em amostra de 80 pupas coletadas em pneus, 30% eram fêmeas. Utilize grau de confiança de 90%.