Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

**NOTAS PARA ACOMPANHAR
AS AULA DA DISCIPLINA**

LME 216
Introdução à Bioestatística Florestal

Prof. Dr. João L.F. Batista

Piracicaba - SP
1997

Sumário

1. Por que Bioestatística Florestal?

2. Tipos de Informações.

3. Resumindo a Informação: "Tabelas"
Variáveis discretas. Definindo as Classes para Variáveis Contínuas. Número de Classes.

4. Resumindo a Informação: "Gráficos"
Análise Gráfica de Dados. Histogramas. Gráficos Ramo-Folha.

5. Resumindo a Informação: "Estatísticas".
Medidas de posição: médias, mediana, moda, quantis/percentis. Propriedades da Média Aritmética.
Medidas de variação: amplitude de variação, variância, desvio padrão, coeficiente de variação. Propriedades da Variância/Desvio Padrão.

6. Gerando Informação: "Amostragem".
Conceitos básicos de amostragem, tipos de amostragem: aleatória simples, conglomerado, sistemática, estratificada

7. Informação e Incerteza: "Probabilidade".
Conceito de probabilidade. Axiomas de probabilidade. Eventos Mutuamente Exclusivos. Probabilidade condicional. Eventos indepedentes. Regra da Probabilidade Total.

8. A Incerteza Domada: "Variáveis Aleatórias Discretas".
Conceito de variável aleatória discreta. Lindado com proporções: a Distribuição Binomial. Lidando com contagens: a Distribuição de Poisson.

9. A Incerteza Domada II: "Variáveis Aleatórias Contínuas".
Conceito de variável aleatória contínua. A Distribuição Uniforme. Distribuição Exponencial.

10. Contornando as Limitações dos Dados: "Distribuições Amostrais"
Distribuições Amostral de Proporções. Distribuição Amostral da Média. Teorema Central do Limite.

11. O Modelo Fundamental: "Distribuição Normal"
Distribuição Normal.

12. Contornando as Limitações dos Dados II: "Mais Distribuições Amostrais"
Distribuição Amostral da Variância: Distribuição de Qui-Quadrado.

13. Tomando Decisões: "Julgando Hipóteses"
Formulando Hipóteses Estatísticas: hipótese nula e hipótese alternativa. Teste de Hipóteses. Erro Tipo I e Erro Tipo II. Nível de Significância e valor-p.

14. Inferência sobre os Parâmetros de uma População: "Intervalo de Confiança"
Distribuição t de Student. Teste t para uma amostra. Intervalo de confiança.

15. Comparando duas Populações: "Teste F"

Razão de Variâncias e Distribuição F. Teste F para duas amostras.

16. Comparando duas Populações II: "Teste t"
Teste t para duas amostras. Teste t pareado e não pareado.

17. Modelo vs. Realidade: "Teste de Qui-Quadrado"
Teste de Qui-quadrado: verificando proporções, verificando a qualidade do ajuste de distribuições.

Apêndices:

A. Respostas de Exercícios Selecionados

B1. Tabelas da Distribuição Normal Padronizada

B2. Tabela da Distribuição t de "Student"

B3. Tabela da Distribuição F

B4. Tabela da Distribuição de Qui-Quadrado

1. Por que Bioestatística Florestal ?

Introdução

Boa parte dos alunos de Engenharia Florestal se pergunta qual o porque da presença de disciplinas de matemática e estatística no seu currículo. Por que eu devo estudar Bioestatística Florestal ? Qual a importância da Bioestatística para o Engenheiro Florestal ? Ao invés de tentarmos uma resposta direta a estas questões analisemos alguns casos típicos da profissão florestal.

Exemplos Profissionais

CASO FLORESTAL 1

Você é diretor de uma fábrica de chapas de madeira (chapas duras). A companhia deseja duplicar a produção da sua fábrica pois o mercado está em franca expansão. Estudando o processo de produção na sua fábrica você chega à conclusão de que o consumo de madeira crescerá em 1.300.000m^3 de madeira/ano com a duplicação. O gerente florestal afirma que as florestas da companhia são capazes de produzir até 1.320.000m^3 de madeira/ano a mais do que vêm produzindo.

1. **Você faz a duplicação da fábrica ? Por que ?**
2. **Qual a diferença na maneira que o seu número (1.300.000 m^3/ano) e o do gerente florestal (1.320.000 m^3/ano) foram gerados ?**

CASO FLORESTAL 2

Uma pesquisadora deseja estimar o tamanho da população de *Furnarios rufus* (João-de-barro) numa dada localidade do interior de Minas Gerais. Num primeiro levantamento, a pesquisadora capturou 31 pássaros e marcou a todos com uma anilha. Num segundo levantamento, ela capturou 47 pássaros dos quais 12 possuíam a anilha referente ao primeiro levantamento.

1. **Qual o tamanho da população de *Furnarios rufus* nessa localidade ?**
2. **O que a pesquisadora deve assumir como verdadeiro para que a partir desses números ela possa chegar a uma estimativa do tamanho da população ?**

CASO FLORESTAL 3

Eucalyptus grandis é uma das espécies arbóreas de maior produtividade quando plantada no Estado de São Paulo. Entretanto, um experimento mostrou que quando a floresta é plantada sem preparo de solo e sem adubação inicial, *Eucalyptus cloeziana* pode alcançar produtividades de 15 a 20% maiores que *E. grandis*.

1. **Como base nessa informação você indicaria *E. cloeziana* em lugar de *E. grandis* para pequenos proprietários rurais que não possuem condições de fazer o preparo de sítio adequado para *E. grandis* ?**
2. **Que informações adicionais são necessárias para tomar uma decisão?**

CASO FLORESTAL 4

Sabe-se que numa floresta tropical não perturbada a abundância de espécies "pioneiras" fica em torno de 10%. Com aumento de perturbações antrópicas a abundância dessas espécies tende a crescer. Na demarcação de uma reserva florestal com área total de 50.000 ha, ecologistas e engenheiros florestais discutem a incorporação de uma área de 7.500 ha onde o levantamento de campo revelou uma abundância de 15% de espécies pioneiras. O objetivo da demarcação da reserva é proteger "áreas representativas de ecossistemas locais que não sofreram influência antrópica significativa".

1. **A área de 7.500 ha deve ou não deve ser incorporada à reserva ?**
2. **Que informações adicionais são necessárias para se tomar uma decisão ?**

Conclusão

- Todo profissional florestal lida com uma grande quantidade de informações qualitativas e quantitativas.
- A tomada de decisão envolve a análise destas informações.
- O raciocínio quantitativo é a base da análise de informações que gera subsíduos à tomada de decisão.
- Esta análise envolve necessariamente um alto grau de incerteza quanto às informações disponíveis, pois estas sempre são incompletas e limitadas.

Conceitos-Chave

INFORMAÇÃO QUANTITATIVA - INFORMAÇÃO QUALITATIVA - TOMADA DE DECISÃO - ANÁLISE DE INFORMAÇÃO - INCERTEZA - RACIOCÍNIO QUANTITATIVO

Exercícios

1.1 Você foi contrato para ser coordenador de uma equipe de profissionais com a missão de realizar um diagnóstico das unidades de conservação no Estado de São Paulo. Que informações de ordem qualitativa você necessitaria.? Que informações de ordem quantitativa seriam necessárias?

1.2 Enumere três justificativas para o Engenheiro Florestal conhecer Bioestatística.

1.3 São pré-requisitos para Bioestatítica Florestal disciplinas de Matemática e Cálculo. Reflita sobre o seu conhecimento nessas disciplinas. Quantifique a proporção dos temas que você realmente domina dentre os tratados nessas disciplinas.

Leitura Sugerida

[WALLIS & ROBERTS, 1964]

Cap. 1 (p.21-38) O Campo da Estatística.
Cap. 2 (p. 39-52) Estatística: Como usá-la devidamente.

2. Tipos de Informações: “Variáveis”

Introdução

O termo “informação” é algo bastante vago na linguagem contidiana. Informação é sinônimo de esclarecimento, notas, argumentos e até fornecimento de dados. Este último sinônimo é o que interessa à Bioestatistica Florestal. Informação deve ser entendida portanto como dados sobre uma realidade ou situação que são fornecidos ou obtidos através de observações ou medições. A Bioestatística Florestal procura manipular as informações (os dados) de modo que elas possam ser mais protamente utilizadas na tomada de decisão. O primeiro passo na manipulação da informação é saber reconhecer os “tipos” básicos que existem.

Escalas Fundamentais

<u>Informação Qualitativa</u>

ESCALA NOMINAL

Conjunto de ATRIBUTOS sem qualquer relação clara entre si.
- Espécies arbóreas numa floresta.
- Procedências de uma espécie
- Classificação dos solos

ESCALA ORDINAL

Conjunto de ATRIBUTOS que se apresentam em uma ORDEM (crescente ou decrescente).
- Julgamentos do tipo: Bom/Médio/Ruim
- Notas de qualidade de fuste das árvores
- Classes de fertilidade dos solos

Informação Quantitativa

ESCALA DE INTERVALO

Os ATRIBUTOS são quantitativos (números) organizados numa escala onde o INTERVALO entre dois valores tem significado real, mas o PONTO ZERO da esca;a é arbitrário.

- ♦ Temperatura
- ♦ Horário do dia
- ♦ Azimute

ESCALA DE RAZÃO

Os ATRIBUTOS são quantitativos (números) organizados numa escala onde tatno o INTERVALO entre dois valores quanto o PONTO ZERO tem significado real.

- ♦ Medidas numéricas em geral
- ♦ Dados de contagem/enumeração
- ♦ Proporções (porcentagens)

PARA REFLETIR

Você pode dizer que uma árvore de 30 metros tem o dobro de altura de uma árvore de 15 metros, ou que um peixe com 9 kg é três vêzes mais pesado que um de 3 kg.

Por que não faz sentido dizer que 40^{o}C é "duas vêzes mais quente" que 20^{o}C ?
Ou que 6:00 horas é um horário "três vêzes mais cedo" que 18:00 hs ?

Exemplo

2.A Classifique as variáveis abaixo de acordo com as quatro escalas fundamentais.

Número de chamadas telefônicas: ________________________

Horário de visitação dos polinizadores de cedro: ________________________

Localização de um formiqueiro numa floresta: ________________________

Tempo de abate de uma árvore: ______________________

Porcentagem de mortalidade de mudas num viveiro: ________________________

Sabores de sorvete na sorveteria Paris: ________________________

Grau de infestação de cupim numa estante: ________________________

Variáveis

VARIÁVEL é um outro termo que utilizamos em Bioestatística para designar informação.

VARIÁVEIS QUALITATIVAS

- Resultam de observações nas escalas Nominal e Ordinal
- Tambem chamadas de **VARIÁVEIS CATEGÓRICAS**
  - Os atributo é representado por *categorias* ou *classes*

VARIÁVEIS QUANTITATIVAS

De representação matemática direta:

X = número de árvores com cancro numa amostra de 10 árvores: $X \in \{0,1,2,3,4,5,6,7,8,9,10\}$

M = taxa de mortalidade das árvores numa floresta nativa: $M \in [0,1]$

A = ângulo de dispersão das sementes de ipê-roxo a partir de uma árvore matrix: $A \in [0,2p\ ]$

Os atributos são representados por escalas numéricas

VARIÁVEL DISCRETA: escala pode ser representada por uma quantidade contável de números (conjunto dos números Naturais):

* X é o número de árvores com cancro numa amostra de 10 árvores:
$$X \in \{0,1,2,3,4,5,6,7,8,9,10\}$$

* Z é a proporção de árvores com cancro numa amostra com 10 árvores:
$$Z \in \{0.0,0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.9,1.0\}$$

* Y é o número plântulas numa área de $2m^2$ do solo de uma floresta nativa:
$$Y \in \{0.1.2.3.4.5.\ldots.\infty\}$$

VARIÁVEL CONTÍNUA: escala é representada por uma quantidade incontável de números (conjunto dos números Reais):

* D é o diâmetro das árvores de uma floresta plantada: $D \in [10,50]$

* W é a biomassa (ton/ha) num ecossistema florestal: $W \in [500,5000]$

*S é a porcentagem saturação de bases num solo florestal: $S \in [0,1]$

Exemplos

2.B Classifique as variáveis abaixo em termos de variáveis categórias e variáveis discretas/contínuas.

Volume de madeira numa floresta. ________________________

Classes de declividade do solo segundo o Código Florestal Brasileiro ________________________

Peso seco das folhas numa árvore. ________________________

Número de árvore mortas em 1 ha de floresta nativa. ________________________

2.C Determine as escalas que melhor representam as variáveis quantitativas da questão anterior.

Variáveis Multidimensionais

Variáveis multidimensionais resultam da combinação de duas ou mais variáveis qualitativas e quantitativas:

* Localização de uma árvore numa parcela de floresta de 500 x 1000 m: $S = (X, Y)$,

onde X e Y são variáveis quantitativas que representam as coordenadas num plano cartesiano:

$$X \in [0{,}500] \; ; \quad Y \in [0{,}1000]$$

* Atributos relacionadas com cada uma das árvores numa amostra: $A = (D, H, S)$,

onde D é o diâmetro da árvore (var. quantitativa contínua), H é a altura (var. quantitativa contínua) e S é a espécie (var. categórica/nominal)

Conceitos-Chave

ESCALAS FUNDAMENTAIS - ESCALA NOMINAL - ESCALA ORDINAL - ESCALA DE INTERVALO - ESCALA DE RAZÃO - VARIÁVEL QUALITATIVA -VARIÁVEL CATEGÓRICA - VARIÁVEL QUANTITATIVA - VARIÁVEL DISCRETA - VARIÁVEL CONTÍNUA

Leitura Essencial

[OLIVEIRA 1977] Cap. 2 (p.5-7) Estatística Descritiva.

[IEMMA, 1992] Cap. I (p.4-8) Noções Introdutórias.

Exercícios

2.1 Classifique as variáveis abaixo segundo as escala fundamentais:

(a) Árvore é classificada como Morta ou Viva: ________________________

(b) Densidade da madeira: ________________________

(c) Formações vegetais brasileiras: ________________________

(d) Crescimento de uma florestas em termos de volume de madeira: ________________________

(e) Formas de vida presente numa floresta nativa: ________________________

(f) Tipos de dormência de sementes: ________________________

(g) Abundância de uma planta herbácea: ________________________

(h) Tipos de deformações numa peça de madeira durante a secagem: ________________________

(i)	Risco de incêndio no dia de hoje numa floresta plantada:	________________________
(j)	Diversidade de espécies arbóreas na bacia do rio Amazonas:	________________________
(k)	Número de calçadodos estudantes da ESALQ:	________________________
(l)	Lista dos 10 livros mais lidos na semana:	________________________
(m)	Figuras de linguagem:	________________________

2.2 Represente as escalas apropriadas para as seguintes variáveis quantitativas:

H = altura das árvores de uma floresta nativa (m);
V = volume de madeira em 1ha de floresta plantada (m^3);
X = número de árvores de mogno numa floresta nativa;
I = posição de inserção dos ramos de uma árvore ao longo da circunferência do tronco (radianos);
Y = número de folhas numa muda de *Eucalyptus camaldulensis*;
S = taxa de sobrevivência das árvores numa floresta nativa;
A = proporção de areia num solo;
E = taxa de autopolinização de uma espécie florestal;
C = Diâmetro das copas das árvores numa floresta de pinus;
D_1 = Declividade do terreno (graus) medida em pontos específicos de uma floresta;
D_2 = Declividade do terreno (%) medida em pontos específicos de uma floresta;
F = face de exposição de uma encosta (graus).

3. Resumindo a Informação: "Tabelas"

Introdução

Talvez a principal tarefa da Bioestatística é a de resumir uma grande quantidade de informação de modo que se torne mas fácil a compreensão dos fenômenos envolvidos e a tomada de decisão. A maneira mais simples de resumirmos a informação contida numa variável quantitativa com um grande número de dados é através de tabelas.

Exemplo 3.A: Diâmetros das Árvores de uma Floresta

Um Engenheiro Florestal mediu o diâmetro de 100 árvores numa floresta nativa.
Os valores obtidos foram:

18.7	11.6	24.9	52.4	33.0	47.6	45.7	11.0	19.8	16.9
12.6	21.9	78.8	13.5	11.9	152.1	22.8	29.0	24.2	16.0
14.0	50.8	20.3	23.5	54.1	30.2	11.8	72.0	13.4	16.3
33.1	15.8	48.3	50.0	55.7	41.1	97.6	34.3	33.6	15.4
22.4	19.6	12.6	19.9	28.1	26.0	37.4	10.8	22.0	25.5
63.2	92.5	18.1	24.6	60.9	83.1	45.2	34.2	20.0	13.1
13.2	15.2	16.2	18.4	12.8	53.1	23.2	15.3	48.7	24.5
10.2	12.0	18.3	20.3	10.6	23.8	22.2	32.2	15.3	14.9
40.0	61.0	53.2	31.7	10.2	52.1	46.9	26.8	17.7	12.8
40.7	13.0	72.4	10.3	31.7	26.3	25.1	12.1	60.7	40.6

O que se pode dizer sobre a floresta com base nestes dados ?

É possível ter uma idéia clara apenas observando os números?

Estes mesmos dados resumidos numa tabela se tornam mais "informativos":

Com uma rápida olhada na tabela já notamos que a maioria das árvores se concentram nas menores classes de diâmetro e que são poucas as árvores grande diâmetro.

Você seria capaz de observar isto nos dados originais ?

CLASSES	FREQÜÊNCIA
10 a 20 cm	40
20 a 30 cm	21
30 a 40 cm	10
40 a 50 cm	10
50 a 60 cm	8
60 a 70 cm	4
70 a 80 cm	3
Acima de 80 cm	4
TOTAL	100

Variáveis Categóricas

As variáveis categóricas podem ser representadas em tabelas "sem perda de informação", pois elas são representadas por classes.

ESPÉCIES ARBÓREAS DE UMA FLORESTA NATIVA DE ACORDO COM A SINDROME DE REGNERAÇÃO

Síndrome de Regeneração	**Número de Espécies**
Heliófitas	7
Oportunistas de Clareira	19
Tolerantes	25
TOTAL	41

SOLOS UTILIZADOS PARA O REFLORESTAMENTO NO ESTADO DE SÃO PAULO

Tipo de Solo	**Freq.**	**Freq. Acumulada**
Muito Arenoso	150	150
Arenoso	143	293
Textura Média	50	343
Argiloso	27	370
TOTAL	370	--

A natureza das variáveis categórias sugere que elas sejam sempre apresentadas na forma tabular.

Variáveis Quantitativas

As variáveis quantitativas podem assumir um grande <u>número de valores</u>

Var. Discretas: número contável de valores

Var. Contínuas: número infinito (não contável) de valores.

Devem ser agrupadas em **CLASSES** antes de serem tabeladas

DEFINIÇÃO DAS CLASSES

1. Número de Classes: $n_{CLASSES} = 1 + 3.33 \log_{10} N_{OBSERVAÇÕES}$ (Algoritmo de Ramsdall)

2. Amplitude dos Dados: $A = X_{MÁXIMO} - X_{MÍNIMO}$

3. Amplitude das Classes: $a = A / n_{CLASSES}$. Arredonde esta amplitude encontrada para um valor "conveniente".

4. Organizar as classes de modo que a primeira classe contenha $X_{MÍNIMO}$ e a última classe contenha $X_{MÁXIMO}$.

Exemplo 3.B: Produção de Resína em Pinus elliottii

Dados brutos de produção de resina (kg) de 40 árvores de *Pinus elliotti*:

```
0.71 2.63 3.63 1.94 3.69 2.77 1.42 2.48 3.77 2.75
2.04 2.16 4.05 1.80 2.22 2.06 1.20 1.67 5.41 1.57
3.09 2.16 3.94 2.06 3.55 3.56 3.57 2.39 2.48 1.53
2.67 2.18 3.93 3.34 2.78 3.26 3.06 3.32 3.37 0.75
```

1) Encontre o número de classes para a construção de uma tabela para estes dados:

$$n_{CLASSES} = 1 + 3.33\log_{10}(40) = 1 + 3.33(1.60206) = 6.33486 \approx 7$$

2) Encontre a amplitude dos dados: $A = 5.41 - 0.71 = 4.70$

3) Encontre a amplitude de classe: $a = 4.70 / 7 = 0.67 \approx 0.7$

4) Defina os limites de cada classe:

Classe	Limite Inferior	Limite Superior
1	0.6	0.6 + 0.7 = 1.3
2	1.3	1.3 + 0.7 = 2.0
3	2.0	2.0 + 0.7 = 2.7
4	2.7	2.7 + 0.7 = 3.4
5	3.4	3.4 + 0.7 = 4.1
6	4.1	4.1 + 0.7 = 4.8
7	4.8	4.8 + 0.7 = 5.5

5) Construa a tabela com a freqüência e a freqüência acumulada.

Classe	Contagem	Freq.	Freq. Acumulada
0.6 - 1.3	///	3	3
1.3 - 2.0	////| /	6	9
2.0 - 2.7	////| ////| //	12	21
2.7 - 3.4	////| ////	9	30
3.4 - 4.1	////| ////	9	39
4.1 - 4.8		0	39
4.8 - 5.5	/	1	40
	TOTAL	40	--

Variáveis Discretas

Algumas variáveis discretas assume um número pequeno de valores.

Neste caso, elas são tabuladas como se fossem variáveis categóricas. Cada valor assumido é tomado como uma classe.

NÚMERO DE BROTOS DEIXADOS EM
CEPAS DE EUCALYPTUS GRANDIS
APÓS O PRIMEIRO CORTE

Dados Brutos: 2 1 2 2 0 3 3 2 1 1 2 2 0 1 1 3 1 2 1 1
2 0 0 3 2 1 2 2 3 0 2 3 3 0 3 2 2 0 1 1

Dados Brutos em ordem: 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 2
2
2 2 2 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3
3

Dados Tabelados:

Número de Brotos	Freq.	Freq. Acumulada
0	7	7
1	11	18
2	14	32

Freqüência Absoluta x Freqüência Relativa

Até agora todas as tabelas apresentadas mostram as Freqüências Absolutas por classe, isto é, o "número" de observações em cada classe.

É comum estarmos interessados não no "número" de observações mas na "proporção" das observações que ocorre numa determinada classe. Esta "proporção" é a freqüência relativa.

PRODUÇÃO DE RESÍNA EM *Pinus elliottii*

Classes (i)	Amplitude	Freqüência Absoluta: Freq. (f_i)	Freqüência Absoluta: Freq. Acumulada (F_i)	Freqüência Relativa: Freq. (p_i)	Freqüência Relativa: Freq. Acumulada (P_i)
1	0.6 - 1.3	3	3	0.075	0.075
2	1.3 - 2.0	6	9	0.150	0.225
3	2.0 - 2.7	12	21	0.300	0.525
4	2.7 - 3.4	9	30	0.225	0.750
5	3.4 - 4.1	9	39	0.225	0.975
6	4.1 - 4.8	0	39	0.000	0.975
7	4.8 - 5.5	1	40	0.025	1.000
	TOTAL	40	--	1.000	--

Relação entre a Freqüência Absoluta e Freqüência Relativa:

$$p_i = f_i \,/\, N_{OBS.}$$

$$P_i = F_i \,/\, N_{OBS.}$$

Conceitos-Chave

TABELA - NÚMERO DE CLASSES - ALGORITMO DE RAMSDALL - AMPLITUDE DOS DADOS - AMPLITUDE DE CLASSE - FREQÜÊNCIA - FREQÜÊNCIA ACUMULADA - FREQÜÊNCIA ABSOLUTA - FREQÜÊNCIA RELATIVA

Leitura Essencial

[IEMMA, 1992] Cap. II: p.20-24. Cap. III: p37-41. Cap. IV: p55-62.

Exercícios

3.1 Construa tabelas (freq. absoluta e relativa) para resumir a informação contida nos seguintes conjuntos de dados:

(a) Número de plântulas num área de 2 x 2 m:

```
4  8 15 18  1  0 17  8  8 16  8  8 20 18  4
7 13 15  1  6  7  9  3 12  4  3  7  8  7  2
```

(b) Altura (cm) de mudas de *Chorisia speciosa* em viveiro:

```
41.5 17.8 27.8 38.7 31.3 36.4 18.9 38.3 27.3 41.0 34.3 30.0 40.2 49.8 26.1 32.6 14.7
43.3 41.5 32.3 30.6 25.0 21.6 38.7 10.3 28.1 26.9 30.0 33.1 28.9    35.1 32.3 33.0 22.2
30.3 34.3 20.2 23.1 27.7 24.9 31.5 29.3 24.5 45.4 38.2 33.9 37.9
```

(c) Densidade da madeira (g/cm^3)de clones de *Eucalyptus grandis:*

```
0.347 0.373 0.297 0.360 0.338 0.357 0.343 0.345 0.392 0.330 0.405 0.364 0.294     0.267
0.413 0.295 0.427 0.333 0.324 0.230 0.445 0.327 0.359 0.446 0.256 0.329 0.471 0.306
0.328
```

(d) Sexo de indivíduos de cotia capturados em armadilhas (M = masculino; F = femenino):

```
F F F F M F F F F F F M F F F M F M F F M M F F F M F F F F F F F M M F M F F F M
```

(e) Classes de capacidade de uso do solo (SR= sem restrições; RCA = com restrições a culturas anuais; P=pastagens; CP=culturas perenes; F=atividades florestais):

```
P   RCA CP  CP  F   P   P   P   CP  CP  SR  CP  P   RCA F   CP  CP  P
F   P   RCA P   RCA SR  CP  SR  P   RCA P   F   CP  RCA SR  F   F   RCA
P   CP  P   CP  RCA P   CP  P   CP  F   SR  F   F   F   F   RCA RCA P
P   CP  CP  CP  CP  CP  RCA CP  CP  RCA CP  SR  CP
```

4. Resumindo a Informação: "Gráficos"

Introdução

Nos dias de hoje, a imprensa utiliza uma grande quantidade de gráficos para apresentar informações ao público. A leitura de revistas como "Veja" e "Isto É" revela que os gráficos são utilizados para apresentar a evolução de índices econômicos e sociais, resultados de pesquisa de opinião e intensão de voto, ou para comparar países, regiões e grupos sociais em termos de atributos qualitativos. Infelizmente, a imprensa nem sempre é imparcial e o tipo de gráfico e a forma de construção podem gerar uma falsa impressão do que os números de fato representam.

Apesar de tais problemas gráfico são poderosas ferramentas para análise de dados. Eles permitem resumir uma grande quantidade de dados em "desenhos espaciais" que revelam rápida e precisamente atributos importantes dos conjuntos de dados representados. As propriedades mais importantes são ressaltadas, sem que a interpretação se perca em meio aos detalhes.

Distribuição de uma Variável

Um aspecto fundamental no estudo de uma variável (informação) é ***como ela varia***, isto é, de que modo ela assume seus diferentes valores.

DISTRIBUIÇÃO

O Padrão de variação de uma variável é chamado de **Distribuição**.

Apresentação da distribuição de uma variável →
- **TABULAR**
- **GRÁFICA**
- **NUMÉRICA**
- **MODELO ESTATÍTICO**

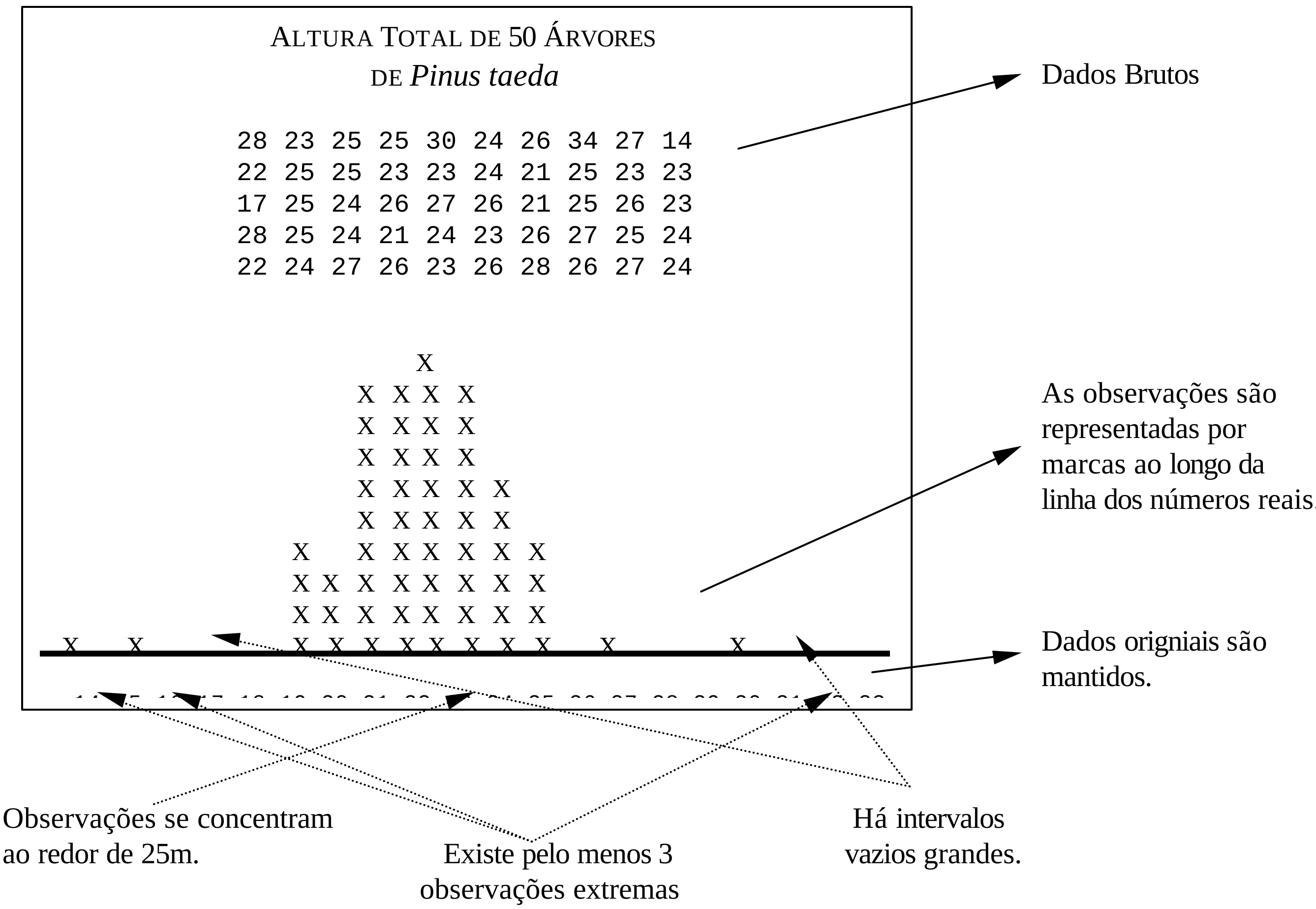
Cada forma de apresentação tem vantagens e desvantagens:

Tabular / Gráfica → Visualização da distribuição e análise exploratória dos dados

Numérica / Modelo Estatístico → Teste de Hipóteses sobre a variável

Gráfico Linha

É o método mais simples de representação gráfica de uma variável.



Observações Extremas (Outliers)

A detecção de intervalos vazios e observações extremas é importante pois nos leva a refletir sobre a qualidade dos dados:

- Por que existe uma disparidade tão grande entre a altura das árvores numa mesma parcela?
- Algum erro de medição ocorreu no campo?
- É razoável prosseguir com as análises dos dados?

OBSERVAÇÕES EXTREMAS (OUTLIERS)

Observações extremas são observações cujos valores estão distintamente abaixo ou acima da maioria das demais observações. Em geral, existem grandes intervalos vazios entre as observações extremas e grupo onde se encontram a maioria dos dados.

Gráfico Ramo-Folha

É outra forma de representar a distribuição de uma variável quantitativa mantendo os valores originais da variável.

Uma **linha vertical** divide os valores das observação numa determinada unidade. No gráfico ao lado a linha representa dezenas de kilogramas (10kg):

4 | 5 → 45kg
10 | 7 → 107 kg

Cada número **à direta** da linha representa uma observação.

PESO (kg) DE ALUNOS DE ENGENHARIA FLORESTAL

Dados Brutos:

```
45  52  53  56  57  58  60  65  65
66  75  53  55  55  58  64  65  66
67  68  68  69  74  74  74  75  75
78  79  79  82 107
```

```
 4 |5
 5 |233556788
 6 |04555667889
 7 |444555899
 8 |2
 9 |
10 |7
```

CONSTRUÇÃO DE UM GRÁFICO RAMO-FOLHA

1) Definir a unidade de medida que dividirá cada valor em duas partes: ramo e folha. No exemplo acima a divisão foi feita em 10kg:

45 kg ==> ramo = 4 / folha = 5
107 kg ==> ramo = 10 / folha = 7

2) Escrever os ramos em ordem crescente verticalmente e passar uma linha vertical à direta deles.

3) Associar cada folha ao respectivo ramo.

4) Ordenar, em cada ramo, as folhas em ordem crescente da direita para esquerda.

Exemplo 4A: Altura de Estudantes de Engenharia Florestal

Construa um gráfico ramo-folha para os dados de altura (m) de estudantes de Engenharia Florestal:

```
1.54 1.55 1.57 1.60 1.60 1.59 1.63 1.63 1.64 1.64 1.72 1.66 1.63 1.67 1.65 1.70
```

```
1.71 1.71 1.71 1.72 1.72 1.71 1.77 1.72 1.71 1.77 1.75 1.77 1.71 1.79 1.75 1.88
```

1) Unidade de media para separação ramo e folha: _________________________

2) Escrever os ramos verticalmente com uma linha à direita deles:

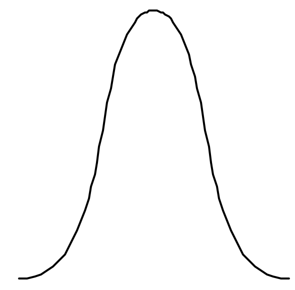
3) Associar cada folha ao respectivo ramo e ordenar as folhas:

4) Analise o gráfico resultante e responda:
   a) Qual a amplitude de variação da altura dos alunos ?
   b) Ocorre observações extremas nestes dados? Quais?
   c) Ocorre diferenciação de grupos de alunos?

Forma da Distribuição de Variáveis

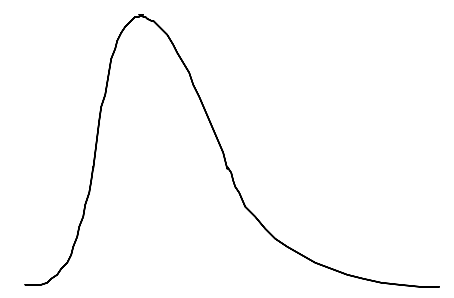
A distribuição de variáveis pode assumir diferentes formas.

Propriedades da distribuição de variáveis → Simetria
→ Unimodal/Bimodal (Número de picos)
→ Uniformidade

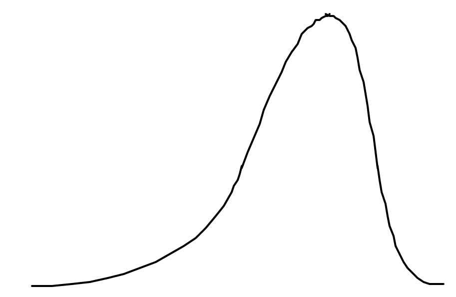


Forma de Sino
Unimoda e Simétrica

Bimodal e
Assimétrica

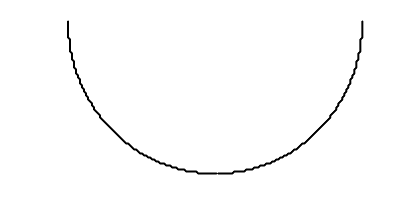


Unmodal e Assimétrica
à Direita



Unmodal e Assimétrica
à Esquerda

Uniforme



Em forma de U

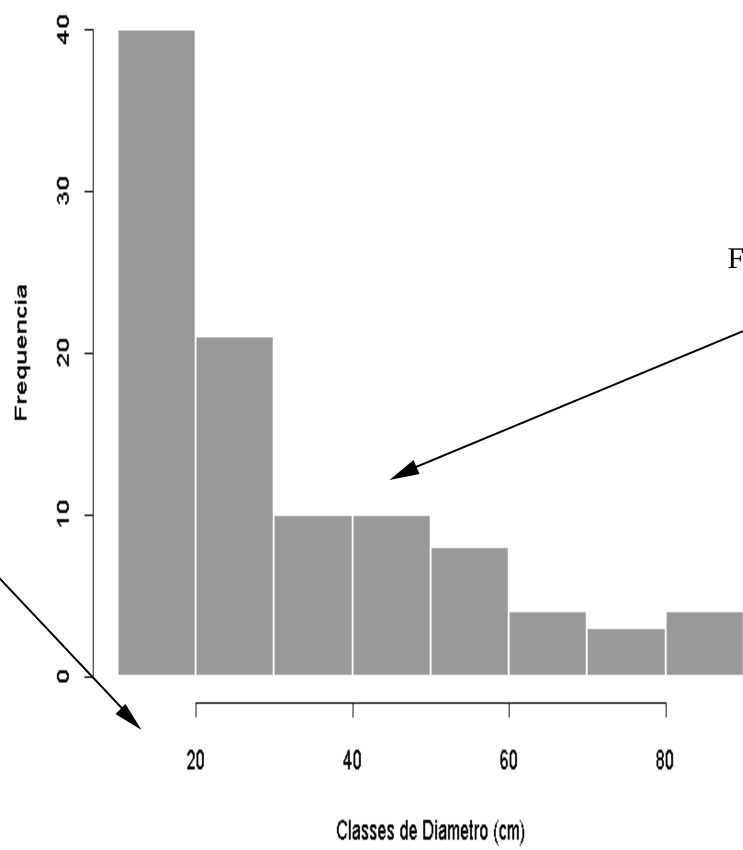
Histogramas

Histogramas são representações gráficas de tabelas de freqüência:

DIÂMETRO (cm) DAS ÁRVORES
NUMA FLORESTA NATIVA

CLASSES	**FREQÜÊNCIA**
10 a 20 cm	40
20 a 30 cm	21
30 a 40 cm	10
40 a 50 cm	10
50 a 60 cm	8
60 a 70 cm	4
70 a 80 cm	3
Acima de 80 cm	4
TOTAL	100

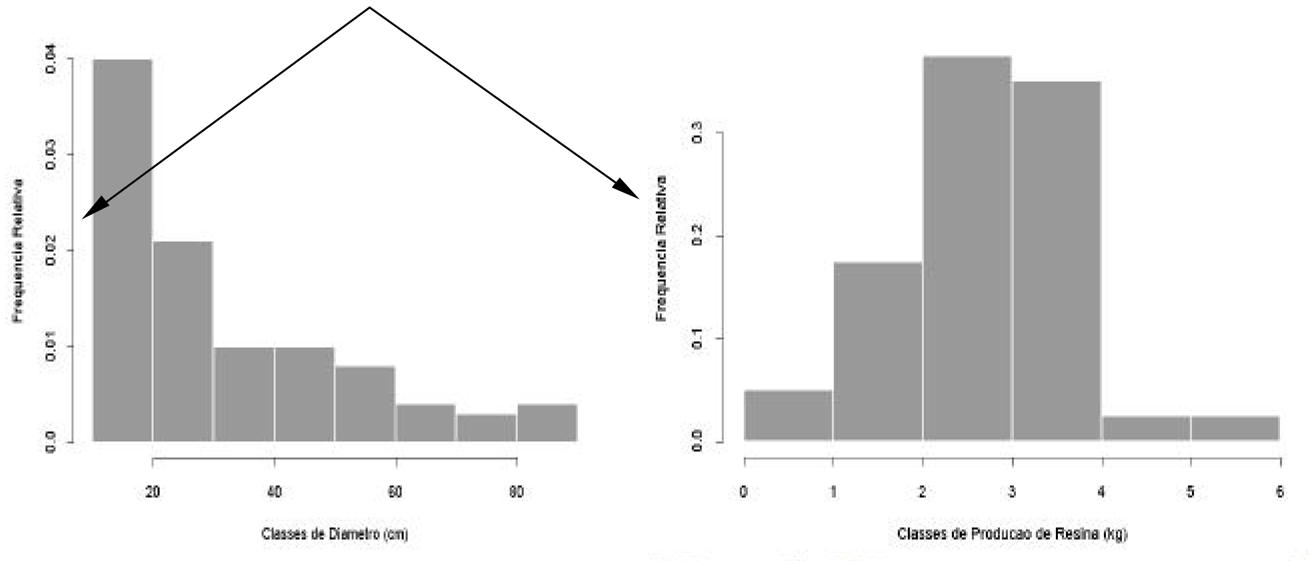
Classes de Diâmetro representadas por pelo eixo-X (abcissas)



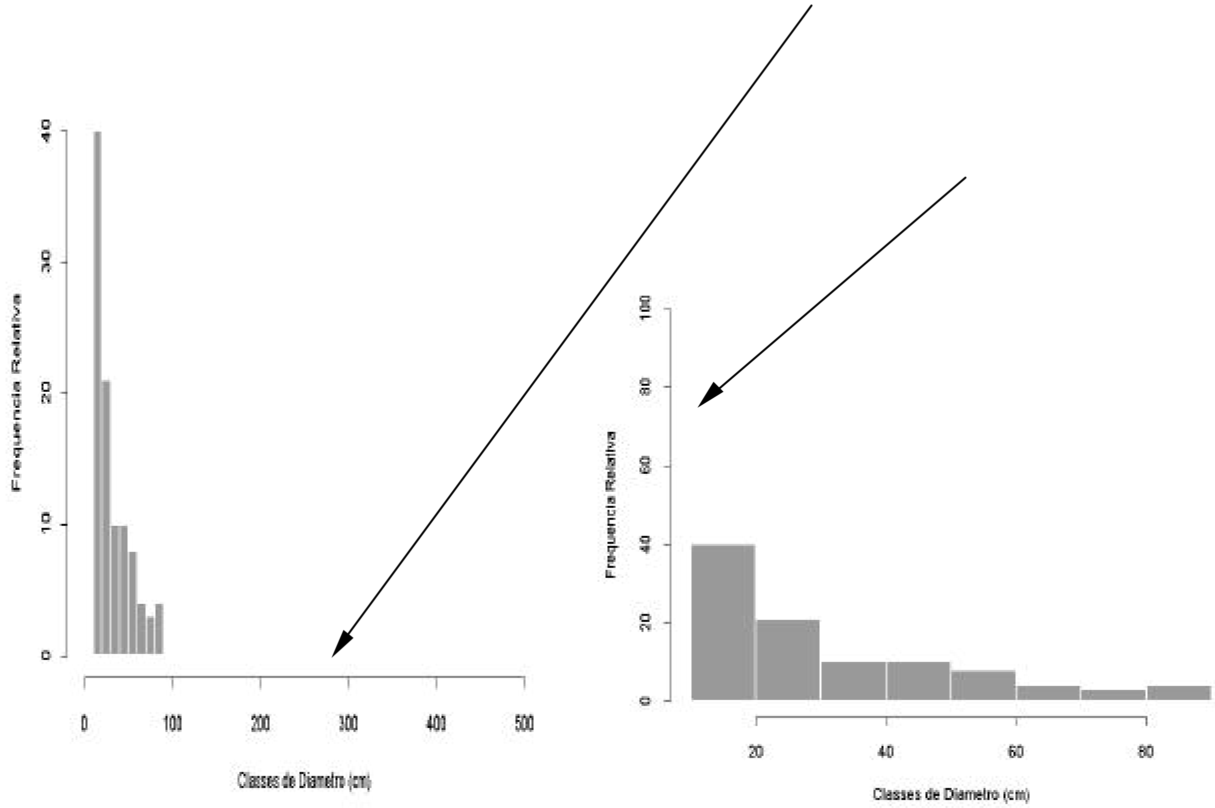
Freqüência em cada classe representada pelo comprimento das barras

ESCALA ADEQUADA DOS EIXOS DO HISTOGRAMA

Eixo-Y (ordenadas) com a Freqüência Relativa facilita a comparação de variáveis com unidades de medidas distintas



Eixos exageradamente grandes em relação à amplitude dos dados geram gráficos "cheios de vazio" que dificultam a observação da forma da distribuição.



Conceitos-Chave

DISTRIBUIÇÃO DE VARIÁVEIS - GRÁFICO RAMO-FOLHA - OBSERVAÇÃO EXTREMA - OUTLIER - FORMA DA DISTRIBUIÇÃO - UNIMODAL/BIMODAL - SIMETRIA - ASSIMETRIA À DIREITA - ASSIMETRIA À ESQUERDA - UNIFORME - HISTOGRAMA

Leitura Essencial

[IEMMA, 1992] Cap. IV: p58-66.

Exercícios

4.1 Os dados abaixo se referem à produção de resina (kg) de árvores de *Pinus elliotti*. Construa um gráfico ramo-folha e analise o comportamento desta variável.

```
0.71 2.63 3.63 1.94 3.69 2.77 1.42 2.48 3.77 2.75
2.04 2.16 4.05 1.80 2.22 2.06 1.20 1.67 5.41 1.57
3.09 2.16 3.94 2.06 3.55 3.56 3.57 2.39 2.48 1.53
2.67 2.18 3.93 3.34 2.78 3.26 3.06 3.32 3.37 0.75
```

4.2 O diâmetro de algumas árvores de duas floresta foi medido num levantamento, obtendo-se os seguintes valores:

FLORESTA A											
16	50	13	8	5	77	93	27	57	28	24	16
49	60	7	5	9	30	8	51	41	33	62	35
9	49	31	107	27	56	26	55	10	18	7	24
17	63	11	34	19	12	40	28	6	19	10	50
16	29	22	10	17	36	42	214	7	10	29	14
12	12	29	76	10	106	52	43	17	16	51	19
21	96	87	29	77	6	9	21	18	6	15	161
32	12	16	29	7	20	37	76	47	6	17	35
30	44	13	56	112	38	15	56	17	34	43	6
52	42	35	25	31	127	9	21	5	154	13	7
FLORESTA B											
38	43	32	18	47	33	38	27	50	34	34	31
28	31	46	27	33	33	38	24	33	23	16	42
22	26	27	32	23	46	30	9	36	47	21	61
34	37	36	30	41	16	7	33	50	11	27	7
23	27	38	23	25	33	30	36	27	32	23	25
30	23	40	15	23	47	35	39	41	46	35	30
42	23	43	35	28	31	35	33	30	30	49	34
48	29	29	30	21	32	28	31	36	22	26	41
33	25	42	30	22	17	40	39	36	55	29	40
42	29	37	29	32	49	17	22	48	31	42	38

a) Descreva a forma da distribuição nas duas florestas.
b) Qual floresta tem maior dispersão no valor dos diâmetros? Por que?
c) Existe observações extremas nesses conjuntos de dados? Quais?
d) Que tipo de florestas são estas (nativa ou plantada)? Por que?

5. Resumindo a Informação: “Estatísticas”

Introdução

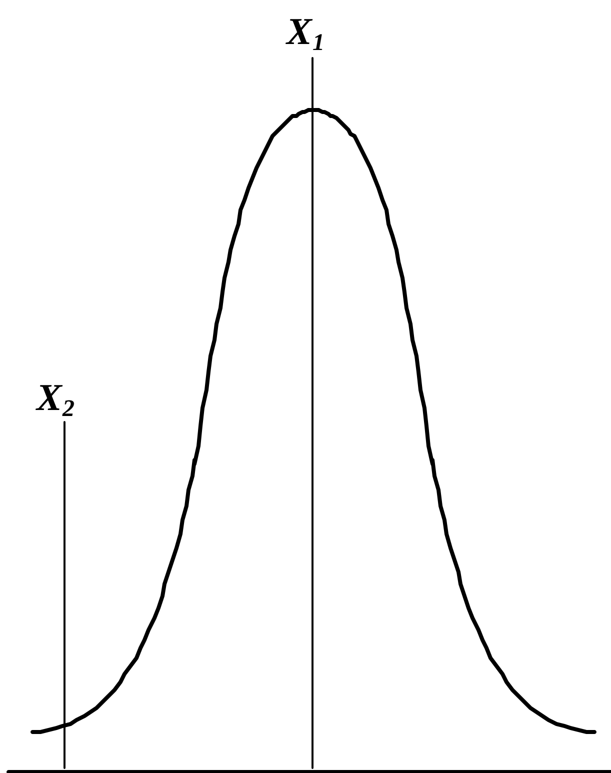
Vimos que a informação contida num conjunto de dados pode ser resumida na forma de tabelas e gráficos. Freqüentemente, entretanto, necessitamos de um “índice” que expresse a certa propriedade dos dados. Por exemplo, numa floresta cada árvore cresce num ritmo particular que depende do seu vigor e da situação ambiental em que está. Como poderíamos representar o crescimento individual “típico” das árvores da floresta ?

ESTATÍSTICAS

As “Estatísticas” são **índices numéricos** que representam propriedades específicas das variáveis

Medidas de “Posição”

A primeira propriedade de uma variável que normalmente estamos interessados se refere a “posições” específicas na distribuição desta variável.



Qual o significado dos valores de X_1 e X_2 na distribuição?

Que locais da distribuição eles representam ?

Média Aritmética

A "média aritmética" (ou simplesmente "média") é a medida de posição mais utilizada pois indica uma "posição central" nos dados.

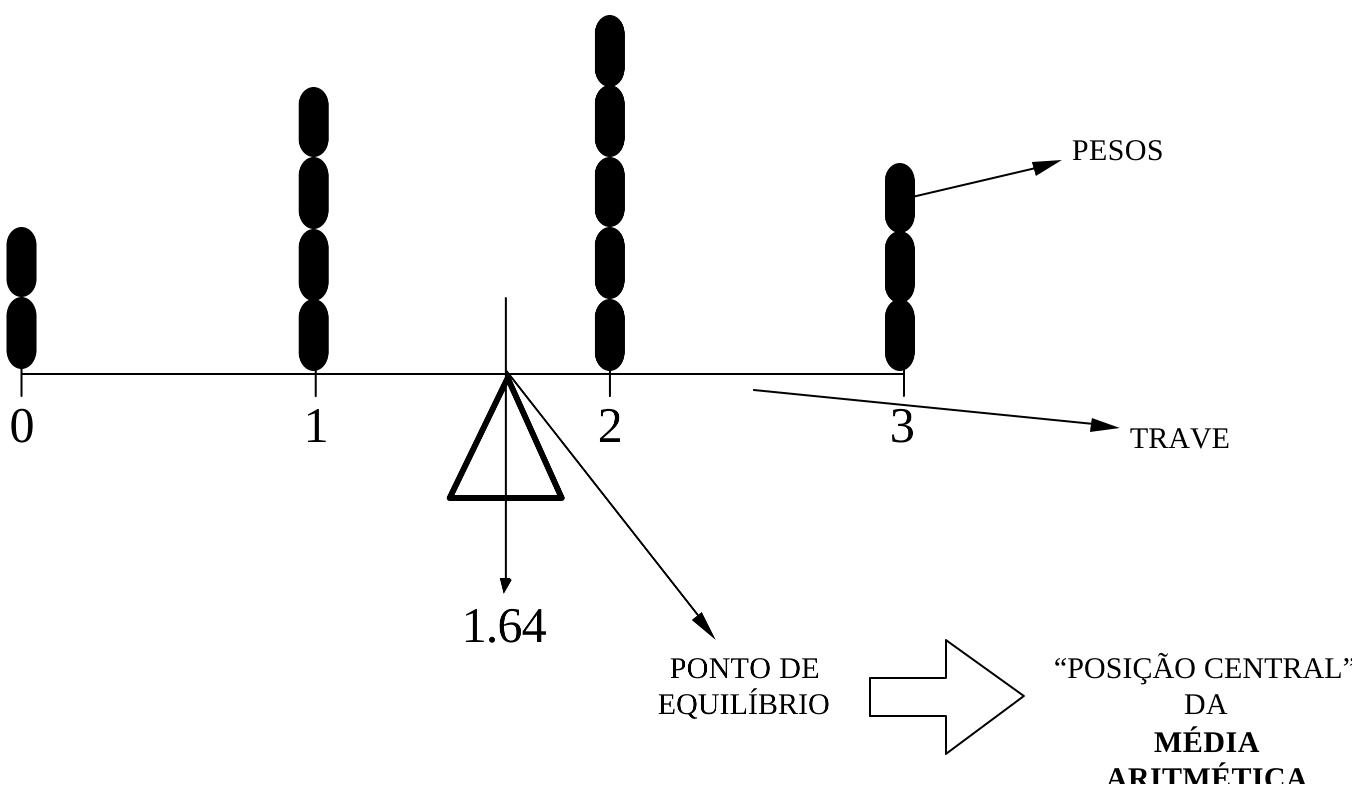
Número de brotações por cepa de *Eucalyptus grandis*: 0, 2, 3, 1, 0, 1, 2, 2, 3, 1, 2, 3, 2, 1.

$$\bar{x} = \frac{0+2+3+1+0+1+2+2+3+1+2+3+2+1}{14} = 1.64$$

O valor $\bar{x} = 1.64$ é o **"número médio"** de brotos por cepa.

- Nenhuma das cepas nos dados originais tinha 1.64 brotos
- É impossível existir uma cepa com 1.64 brotos

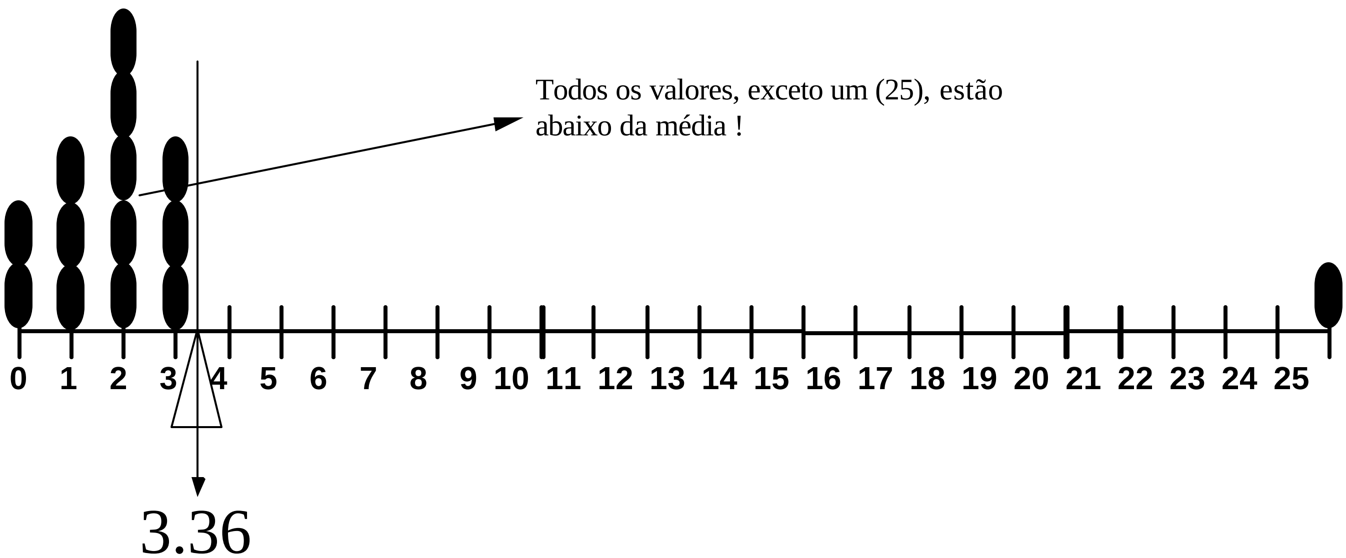
<u>O que significa um número médio de 1.64 brotos por cepa ?</u>



<u>A média aritmética nem sempre está no "CENTRO"</u>

Número de brotações por cepa de *Eucalyptus saligna*: 0, 2, 3, 1, 0, 1, 2, 2, 3, 1, 2, 3, 2, 25.

$$\bar{x} = \frac{0+2+3+1+0+1+2+2+3+1+2+3+2+25}{14} = 3.36$$



Somatórios e suas Propriedades

A média é obtida através do "somatório" dos valores observados.

Se X é uma variável quantitativa, então, sua média será:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + x_4 + x_5 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

PROPRIEDADES:

<u>Somatório de n observações da variável X:</u>

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + x_3 + x_4 + x_5 + \ldots + x_n$$

Somatório de n observações de uma constante k:

$$\sum_{i=1}^{n} \mathrm{k} = \underbrace{\mathrm{k} + \mathrm{k} + \mathrm{k} + \mathrm{k} + \mathrm{k} + \ldots + \mathrm{k}}_{n \text{ vêzes}} = n\mathrm{k}$$

Somatório de n observações de $k\,X$:

$$\sum_{i=1}^{n} \mathrm{k}x_i = \mathrm{k}\ \mathrm{x}_1 + \mathrm{k}\ \mathrm{x}_2 + \mathrm{k}\ \mathrm{x}_3 + \mathrm{k}\ \mathrm{x}_4 + \ldots + \mathrm{k}\ \mathrm{x}_\mathrm{n} =$$

$$= \mathrm{k}\ (\mathrm{x}_1 + \mathrm{x}_2 + \mathrm{x}_3 + \mathrm{x}_4 + \ \ldots\ + \mathrm{x}_\mathrm{n}) = \mathrm{k}\sum_{i=1}^{n} x_i$$

Somatório de n observações de $(X - Y)$:

$$\sum_{i=1}^{n} (x_i - y_i) = (\mathrm{x}_1 - \mathrm{y}_1) + (\mathrm{x}_2 - \mathrm{y}_2) + (\mathrm{x}_3 - \mathrm{y}_3) + \ \ldots\ + (\mathrm{x}_\mathrm{n} - \mathrm{y}_\mathrm{n}) =$$

$$= (\mathrm{x}_1 + \mathrm{x}_2 + \mathrm{x}_3 + \ \ldots\ + \mathrm{x}_\mathrm{n}) - (\mathrm{y}_1 + \mathrm{y}_2 + \mathrm{y}_3 + \ \ldots\ + \mathrm{y}_\mathrm{n}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i$$

Somatório de n observações de $(X - k)$:

$$\sum_{i=1}^{n} (x_i - \mathrm{k}) = (\mathrm{x}_1 - \mathrm{k}) + (\mathrm{x}_2 - \mathrm{k}) + (\mathrm{x}_3 - \mathrm{k}) + \ \ldots\ + (\mathrm{x}_\mathrm{n} - \mathrm{k}) =$$

$$= (\mathrm{x}_1 + \mathrm{x}_2 + \mathrm{x}_3 + \ \ldots\ + \mathrm{x}_\mathrm{n}) - (\mathrm{k} + \mathrm{k} + \mathrm{k} + \ \ldots\ + \mathrm{k}) = \sum_{i=1}^{n} x_i - n\mathrm{k}$$

Somatório de n observações da variável X^2 (**soma de quadrados**):

$$\sum_{i=1}^{n} x_\mathrm{i}^2 = \mathrm{x}_1^2 + \mathrm{x}_2^2 + \mathrm{x}_3^2 + \ \ldots\ + \mathrm{x}_\mathrm{n}^2$$

Quadrado do Somatório de n observações da variável X (**quadrado da soma**):

$$\left(\sum_{i=1}^{n} x_i\right)^2 = (x_1 + x_2 + x_3 + \ldots + x_n)^2$$

Somatório de n observações de $(X - k)^2$:

$$\sum_{i=1}^{n}(x_i - \mathrm{k})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\mathrm{k} + \mathrm{k}^2) = \sum_{i=1}^{n} x_i^2 - 2\mathrm{k}\sum_{i=1}^{n} x_i - n\mathrm{k}^2$$

Somatório do produto de n observações de X e Y: (**soma de produtos**):

$$\sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + x_3 y_3 + \ldots + x_n y_n$$

Exemplo 5.A: Brotos de cepas de Eucalipto

Aplique as propriedades dos somatórios nas seguintes informações:

X é número de brotos por cepa de Eucalipto: `{3, 4, 1, 4, 3, 3, 3, 2};`
Y é altura das cepas: `{10.1, 11.1, 10.7, 13.1, 14.5, 13.5, 12.5};`
n é o número de observações de X e Y;
k = 5.

Somatório de n observações da variável X:

Somatório de n observações de uma constante k:

Somatório de n observações de $k\ X$:

Somatório de n observações de $(X - Y)$:

Somatório de n observações de $(X - k)$:

Soma de Quadrados de n observações de X:

Quadrado da Soma de n observações de X:

<u>Soma de Quadrados de n observações de $(X - k)$</u>:

<u>Soma de Produtos de n observações de X e Y</u>:

Cálculo da Média para Dados Tabelados

Quando dos dados estão agrupados a média é obtida pelo
soma do produto do "centro de classe" e da freqüência em cada classe.

$$\overline{x} = \frac{\sum_{i=1}^{k} f_i x_i}{\sum_{i=1}^{k} f_i} = \frac{f_1 x_1 + f_2 x_2 + f_3 x_3 + \ldots + f_n x_n}{f_1 + f_2 + f_3 + \ldots + f_n}$$

Exemplo 5.B: Qual o Diâmetro Médio numa Floresta Nativa ?

Distribuição dos Diâmetros
numa Floresta Nativa

CLASSES	CENTRO DE CLASSE (xi)	FREQÜÊNCIA (fi)
10 a 20 cm	15 cm	40
20 a 30 cm	25 cm	21
30 a 40 cm	35 cm	10
40 a 50 cm	45 cm	10
50 a 60 cm	55 cm	8
60 a 70 cm	65 cm	4
70 a 80 cm	75 cm	3
80 a 90 cm	85 cm	4
TOTAL		100

$$\overline{D} = \frac{15(40) + 25(21) + 35(10) + 45(10) + 55(8) + 65(4) + 75(3) + 85(4)}{100}$$

$$\overline{D} =$$

Mediana

MEDIANA

Divide as observações em dois grupos de igual tamanho, isto é, 50% das observações estão abaixo da mediana e 50% estão acima.

Número de brotações por cepa de *Eucalyptus grandis*:

Dados brutos: `0, 2, 3, 1, 0, 1, 2, 2, 3, 1, 2, 3, 2, 1.`

Dados em ordem: `0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3,`

MEDIANA ENTRE 2 E 2
MEDIANA = 2

Número de brotações por cepa de *Eucalyptus saligna*:

Dados brutos: `0, 2, 3, 1, 0, 1, 2, 2, 3, 1, 2, 3, 2, 25`

Dados em ordem: `0, 0, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 25`

MEDIANA ENTRE 2 E 2
MEDIANA = 2

QUAL O EFEITO DE OBSERVAÇÕES EXTREMAS SOBRE A MÉDIA E A MEDIANA?

Fórmula de Cáculo da Mediana (x_{MED}):

- Se n é impar: $x_{MED} = x_{[(n+1)/2]}$

Exemplo: {19, 20, 23, 23, 24, 25, 32, 36, 39, 40, 41} $\Longrightarrow x_{MED} = x_{[6]} = 25$

- Se n é par: $x_{MED} = \dfrac{x_{[n/2]} + x_{[(n/2)+1]}}{2}$

Exemplo: {19, 20, 23, 23, 24, 25, 32, 36, 39, 40, 41, 44} $\Rightarrow x_{MED} = \dfrac{x_{[6]} + x_{[7]}}{2} = \dfrac{25+32}{2} = 28.5$

Moda

MODA

É o valor que ocorre com maior freqüência nas observações de uma variável.

Número de brotações por cepa de *Eucalyptus grandis*:

Dados brutos:

```
0, 2, 3, 1, 0, 1, 2, 2, 3, 1, 2, 3, 2, 1.
```

Dados em ordem:

```
0, 0, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3.
```

MOD

Número de Brotos	Freqüência
0	2
1	4
2	**5**
3	3
TOTAL	14

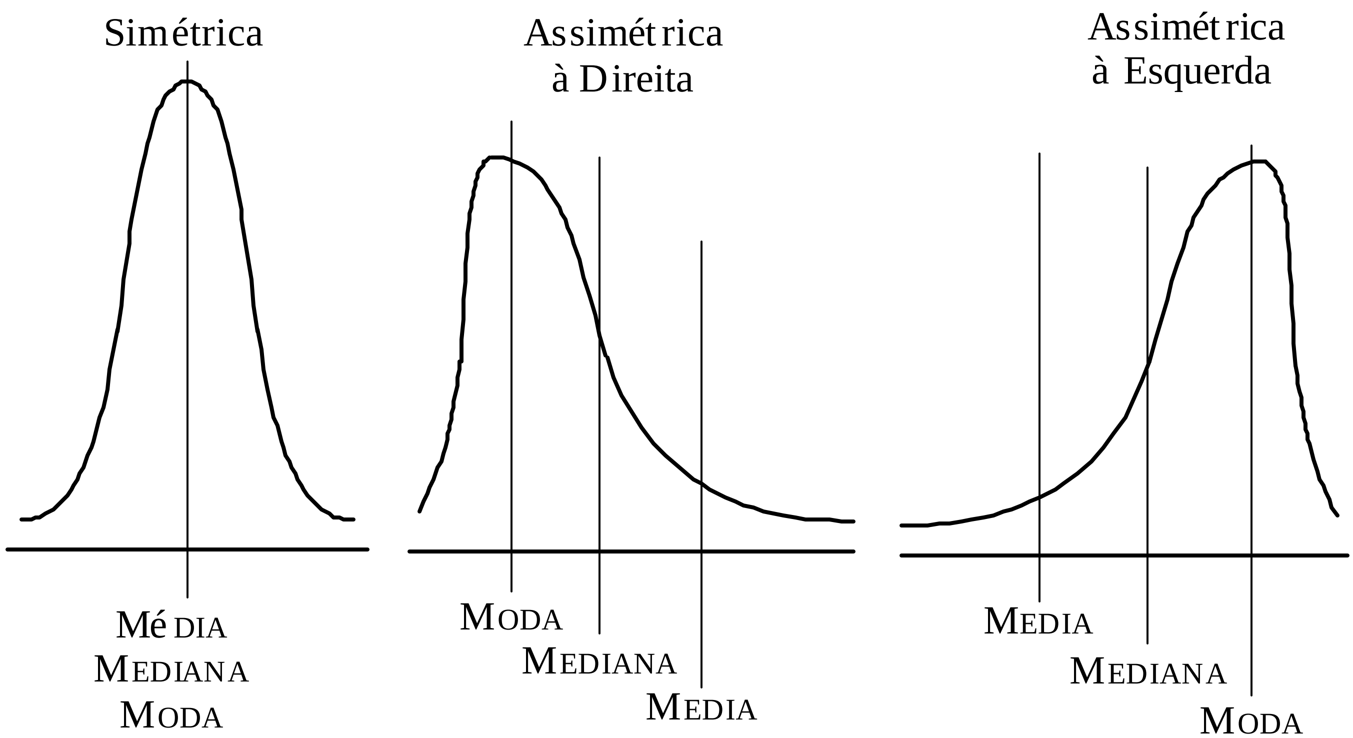
ESPÉCIES ARBÓREAS DE UMA FLORESTA NATIVA DE ACORDO COM A SINDROME DE REGNERAÇÃO

Síndrome de Regeneração	Número de Espécies
Heliófitas	7
Oportunistas de Clareira	19
Tolerantes	**25**
TOTAL	41

Nas Variáveis Categóricas a moda é a classe ou categoria de maior freqüência, sendo a "Estatística" de posição mais apropriada para esse tipo de variável.

MODA

Distribuição dos Dados versus Média - Mediana - Moda



Quartis

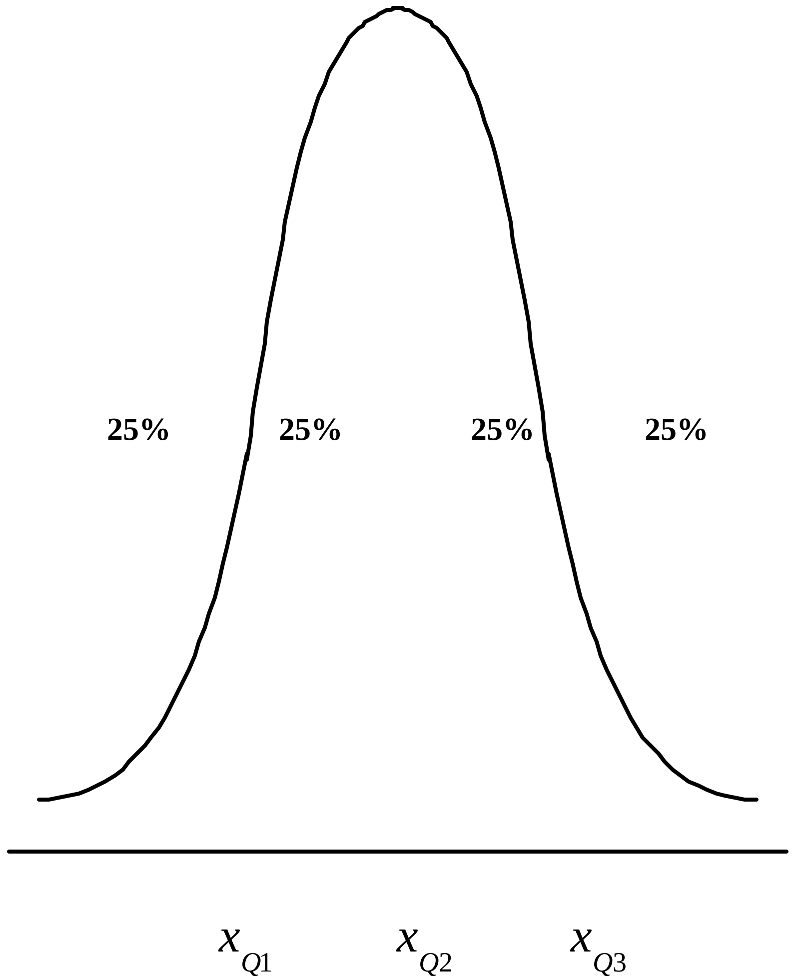
QUARTIS

Os Quartis são as posições na distribuição de uma variável que a divide em "quartos".

PRIMEIRO QUARTIL (Q1): 25% das observações são menores que ele.

SEGUNDO QUARTIL (Q2): 50% das observações estão abaixo dele (mediana).

TERCEIRO QUARTIL (Q3): 75% das observações estão abaixo dele.



Embora possamos calcular os quartis diretamente a partir das observações, é mais prático realizar o cálculo a partir de dados agrupados (Tabela de Freqüência).

$$x_Q = l_k + \left[\frac{\left(F_Q - F_{k-1}\right)}{f_k} \right] a_k$$

x_Q = quartil (1o., 2o. ou 3o.);
k = classe quartil, isto é, a classe que contém o quartil;
l_k = limite inferior da classe quartil;
F_Q = freqüência acumulada até o quartil:

$$F_{Q1} = \frac{n}{4}; \qquad F_{Q2} = \frac{n}{2}; \qquad F_{Q3} = \frac{3n}{4}$$

F_{k-1} = freqüência acumulada até a classe anterior à classe quartil;
f_k = freqüência da classe quartil;
a_k = amplitude da classe quartil.

Exemplo 5.C: Quartis da Distribuição de Diâmetros.

$$F_{Q1} = 100/4 = 25 \qquad F_{Q2} = 100/2 = 50 \qquad F_{Q3} = (3)(100)/4 = 75$$

DISTRIBUIÇÃO DOS DIÂMETROS
NUMA FLORESTA NATIVA

CLASSES	CENTRO DE CLASSE (xi)	FREQÜÊNCIA (fi)	FREQÜÊNCIA ACUMULADA (Fi)
10 a 20 cm	**15 cm**	**40**	**40**
20 a 30 cm	**25 cm**	**21**	**61**
30 a 40 cm	35 cm	10	71
40 a 50 cm	**45 cm**	**10**	**81**
50 a 60 cm	55 cm	8	89
60 a 70 cm	65 cm	4	93
70 a 80 cm	75 cm	3	96
80 a 90 cm	85 cm	4	100
TOTAL	--	100	--

CLASSES QUARTIS

Q1

Q2

Q3

$$x_{Q1} = l_1 + \left[\frac{\left(F_{Q1} - F_0\right)}{f_1} \right] a_1 =$$

$$x_{Q2} = l_2 + \left[\frac{\left(F_{Q2} - F_1\right)}{f_2} \right] a_2 =$$

$$x_{Q3} = l_4 + \left[\frac{\left(F_{Q3} - F_3\right)}{f_4} \right] a_4 =$$

Percentis

PERCENTIS

Os **Percentis** são uma generalização dos Quartis. Na distribuição de uma variável quantitativa, o percentil representa a posição abaixo da qual está uma dada ***porcentagem*** dos dados.

Por exemplo:

o percentil 13% é o valor abaixo do qual estão 13% das observações;

o percentil 64.9% é o valor abaixo do qual estão 64.9% das observações;

A fórmula de cálculo dos percentis (dados tabelados) é análoga à fórmula dos quartis:

$$x_P = l_k + \left[\frac{\left(F_P - F_{k-1} \right)}{f_k} \right] a_k$$

x_P = percentil desejado;

k = classe percentil, isto é, a classe que contém o percentil;

l_k = limite inferior da classe percentil;

$F_P = (P/100)n$ = freqüência acumulada até o percentil, sendo P o percentual correspondente ao percentil:

F_{k-1} = freqüência acumulada até a classe anterior à classe percentil;

f_k = freqüência da classe percentil;

a_k = amplitude da classe percentil.

Exemplo 5.D: Desbaste numa Floresta Plantada

Um Engenherio Floresta deseja fazer um desbaste numa floresta plantada, retirando 25% das árvores das menores árvores. Qual o diâmetro abaixo do qual todas árvores serão removidas?

$$F_{25\%} = \left(\frac{25}{100} \right) 196 = 49$$

DISTRIBUIÇÃO DOS DIÂMETROS
NUMA FLORESTA NATIVA

CLASSES	CENTRO DE CLASSE (xi)	FREQÜÊNCIA (fi)	FREQÜÊNCIA ACUMULADA (Fi)
5 a 10 cm	7.5 cm	17	17
10 a 15 cm	12.5 cm	21	38
15 a 20 cm	**17.5 cm**	**35**	**73**
20 a 25 cm	22.5 cm	40	113
25 a 30 cm	27.5 cm	38	151
30 a 35 cm	32.5 cm	24	175
35 a 40 cm	37.5 cm	13	188

CLASSE PERCENTIL

P=25%

$$x_{25\%} = l_3 + \left[\frac{\left(F_{25\%} - F_2 \right)}{f_3} \right] a_3 =$$

PARA REFLETIR

Uma Engenheira Florestal deseja construir uma casa de madeira. Sua opção é utilizar madeira tratada devido ao menor custo em relação à madeira de essencias nativas. Ela consultou três fornecedores de madeira, que apresentaram preços semelhantes, mas as garantias da vida útil da madeira tratada foi bastante distinta:

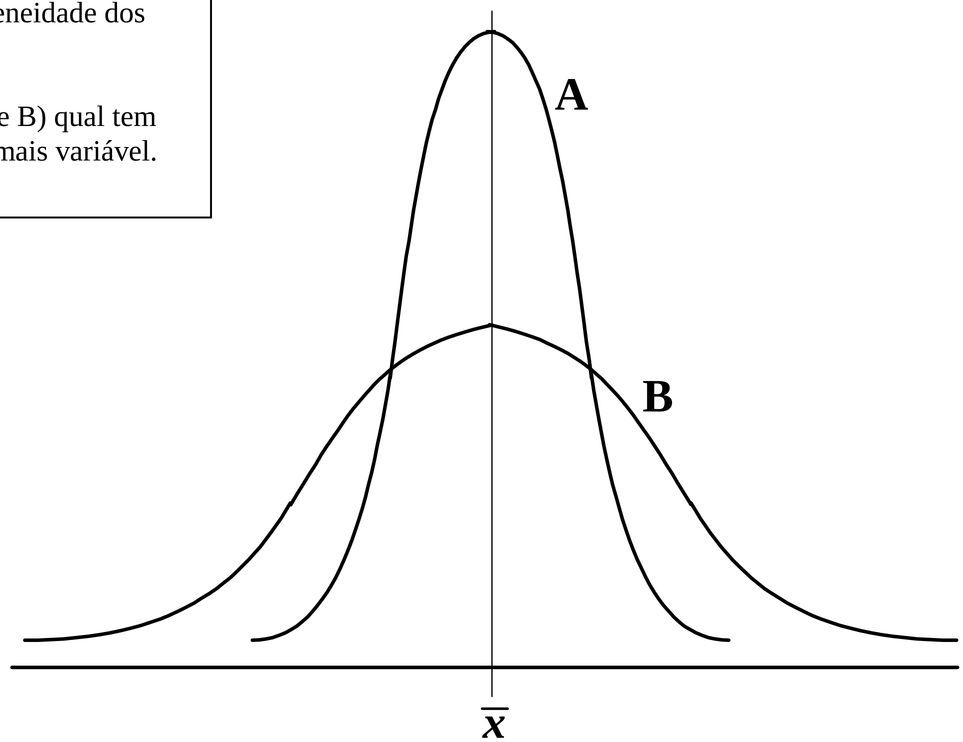
Fornecedor A: vida útil média é 25 anos.
Fornecedor B: vida útil mediana é 25 anos.
Fornecedor C: percentil 70% da vida útil é 25 anos.

De qual fornecedor ela deve adquirir a madeira? Por que?

Medidas de "Dispersão"

Uma "Estatística" de "dispersão" se refere à variabilidade ou heterogeneidade dos dados.

Nas duas distribuições (A e B) qual tem maior dispersão ? Qual é mais variável.



Exemplo 5.E: Controle de Qualidade

Uma Empresa de Papel realizou testes para comparar três máquinas de produção de papel. Em cada máquina foram tomadas 5 amostras conforme a figura ao lado.

Qual das máquinas a empresa deve adquirir? Por que?

GRAMATURA DO PAPEL PRODUZIDO EM DIFERENTES MÁQUINAS

Amostra		**Máquinas**	
	A	**B**	**C**
1	200	152	205
2	210	248	203
3	190	260	195
4	215	200	197
5	185	140	200
Média	200	200	200

Amplitude de Variação e Distância Interquartil

A maneira mais simples de medir a variabilidade de uma variável é através da "distância" entre duas posições na distribuição.

Amplitude de Variação: $A = x_{MÁXIMO} - x_{MÍNIMO}$

Distância Interquartil: $\mathrm{DIQ} = x_{Q3} - x_{Q1}$

Variância

Outra forma de se medir a variabilidade de uma variável é quantificando a dispersão das observações em relação a um ponto específico na distribuição.

Ponto Específico na Distribuição

MÉDIA: $\overline{x}$

Distância das Observações para a Média

DESVIO: $\left(x_i - \overline{x}\right)$

ALTURA DE CEPAS DE *Eucalyptus grandis*

Observações	Desvio	Quadrado do Desvio
1	(10.1 - 12.21) = -2.11	$(10.1 - 12.21)^2$ = 4.45
2	(11.1 - 12.21) = -1.11	$(11.1 - 12.21)^2$ = 1.23
3	(10.7 - 12.21) = -1.51	$(10.7 - 12.21)^2$ = 2.28
4	(13.1 - 12.21) = 0.89	$(13.1 - 12.21)^2$ = 0.79
5	(14.5 - 12.21) = 2.29	$(14.5 - 12.21)^2$ = 5.24
6	(13.5 - 12.21) = 1.29	$(13.5 - 12.21)^2$ = 1.66
7	(12.5 - 12.21) = 0.29	$(12.5 - 12.21)^2$ = 0.08
SOMA	**0.03**	**15.75**

Soma dos Desvios é sempre **zero** (exceto por problemas de arredondamento)

Melhor utilizar a SOMA DE QUADRADOS DOS DESVIOS que será sempre **positiva**.

VARIÂNCIA

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

é o "**Desvio Quadrado Médio**", mede a variabilidade independentemente do número de observações (n).

SOMA DE QUADRADOS

$$SQ = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

será maior quanto maior o número de observações (n)

Desvio Padrão

Como as variáveis que trabalhamos na Engenharia Florestal possuem **UNIDADES DE MEDIDA**, é importante considerá-las quando medimos a heterogeneidade dos dados.

Variável	**Unidade de Medida**	**Unidade da Variância**
altura de cepas	cm	**cm^2**
peso seco de folhas	kg	**kg^2**
produção de madeira	m^3/ha	**m^6/ha^2**
concentração de N	mg/cm^3	**mg^2/cm^6**
concentração de Z	ppm	**ppm^2**
pressão de pneu	lb/pol^2	**lb^2/pol^4**
resistência a compressão	kg/cm^2	**kg^2/cm4**
idade de corte	ano	**ano^2**
área basal	m^2/ha	**m^4/ha^2**

A variância sempre elava ao quadrado as unidades de medida, gerando escalas sem sentido prático

Se utilizarmos a raiz quadrada da variância, recuperaremos as unidades originais.

DESVIO PADRÃO

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

mede a variabilidade independentemente do número de observações (n) e com a mesma unidade de medida da média.

Fórmula Prática para Cálculo
(x_i = observação individual)

SOMA DE QUADRADOS

$$\text{SQ} = \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 = \sum_{i=1}^{n} x_i^2 - \frac{\left(\sum_{i=1}^{n} x_i\right)^2}{n}$$

VARIÂNCIA

$$s^2 = \frac{\text{SQ}}{n-1}$$

DESVIO PADRÃO

$$s = \sqrt{s^2}$$

Fórmula para Dados Tabelados
(x_i = centro de classe)

SOMA DE QUADRADOS

$$\text{SQ} = \sum_{i=1}^{k} f_i\left(x_i - \bar{x}\right)^2$$

VARIÂNCIA

$$s^2 = \frac{\text{SQ}}{\left(\sum_{i=1}^{k} f_i\right) - 1}$$

DESVIO PADRÃO

$$s = \sqrt{s^2}$$

Exemplo 5.F: Redimento de Equipes de Trabalho

Duas equipes de medição foram analisadas em termos do número de parcelas medidas por dia:

Equipe A: `{24, 16, 26, 10, 15, 12, 13, 24, 18, 12}`

Equipe B: `{19, 16, 16, 18, 18, 19, 20, 17, 17, 15}`

- Qual equipe é mais produtiva?
- Qual equipe é mais consistente?

<u>Média:</u>

<u>Soma de Quadrados:</u>

<u>Variância:</u>

<u>Desvio Padrão:</u>

Exemplo 5.G: Estrutura de uma Floresta Plantada

A distribuição dos diâmetros das árvores de uma floresta plantada é apresentada na tabela abaixo.

Classes (cm)	Centro de Classe (x_i)	Freqüência (f_i)	$f_i x_i$	$f_i(x_i - \overline{x})$	$f_i(x_i - \overline{x})^2$
4 - 6	5	17	17(5) =	17(5-) =	17(5-)^2 =
6 - 8	7	21	21(7) =	21(7-) =	21(7-)^2 =
8 - 10	9	35	35(9) =	35(9-) =	35(9-)^2 =
10 - 12	11	40	40(11) =	40(11-) =	40(11-)^2 =
12 - 14	13	38	38(13) =	38(13-) =	38(13-)^2 =
14 - 16	15	24	24(15) =	24(15-) =	24(15-)^2 =
16 - 18	17	13	13(17) =	13(17-) =	13(17-)^2 =
18 - 20	19	8	8(19) =	8(19-) =	8(19-)^2 =
	TOTAL	196			

Qual o diâmetro médio das árvores nessa floresta ?

Qual é a soma de quadrados ?

Qual é a variância?

Qual é o desvio padrão?

Coeficiente de Variação

O Coeficiente de variação é uma forma de se medir a variabilidade de uma variável de modo independente da UNIDADE DE MEDIDA utilizada ou da ORDEM DE GRANDEZA dos dados.

COEFICIENTE DE VARIAÇÃO

$$CV = \frac{s}{\overline{x}} 100$$

mede a variabilidade numa escala percentual, independe da unidade de medida ou da ordem de grandeza da variável.

Razão entre desvio padrão e média torna o CV um número puro

Exemplo 5.H: Comparando a Dispersão de Diferentes Variáveis

Variável	Observações	Média	Desvio Padrão	CV (%)
Altura de árvores (m)	{ 15.2 , 20.3 , 35.2 , 34.1 , 27.2 }	26.4 m	8.66 m	32.8
Altura de arbustos (m)	{ 2.1 , 1.0 , 3.5 , 5.2 , 4.7 }	3.3 m	1.76 m	53.2
Diâmetro de árvores (cm)	{ 14.8 , 21.2 , 36.1 , 35.0 , 26.9 }	26.8cm	9.07cm	33.8

Transformação Linear

X é o número de brotações por cepa de *Eucalyptus grandis:*

x = **{ 0, 2, 3, 1, 0, 1, 2, 2, 3, 1, 2, 3, 2, 1}**

$\boxed{\bar{x} = 1.64}$

$\boxed{s_X = 1.01}$

$Y = X + 5$ (soma-se 5 a todas as observações de X):

y = **{ 5, 7, 8, 6, 5, 6, 7, 7, 8, 6, 7, 8, 7, 6}**

$\boxed{\bar{y} = \quad}$

$\boxed{s_Y = \quad}$

Como seria a média e desvio padrão da variável $W = X - 1$?

$Z = 2X$ (multiplica-se por 2 todas as observações de X):

z = **{ 0, 4, 6, 2, 0, 2, 4, 4, 6, 2, 4, 6, 4, 2}**

$\boxed{\bar{z} = \quad}$

$\boxed{s_Z = \quad}$

Como seria a média e desvio padrão da variável $T = -7X$

TRANSFORMAÇÃO LINEAR

Se X representa os valores originais com média $\bar{x}$ e desvio padrão s_X, chamamos de transformação linear de X os valores: $Y = aX + b$.

Os novos valores Y terão: média: $\bar{y} = a\bar{x} + b$

desvio padrão: $s_Y = |a| s_X$

Padronização

Façamos uma transformação expecial nos dados originais de número de brotos por cepa:

	Dados Originais ($\boldsymbol{x}$)	$x - \bar{x}$	$\frac{x-\bar{x}}{s_X} = \left(\frac{1}{s_X}\right)x + \left(-\frac{\bar{x}}{s_X}\right) \equiv ax + b$
	0		
	2		
	3		
	1		
	0		
	1		
	2		
	2		
	3		
	1		
	2		
	3		
	2		
	1		
MÉDIA	1.64		
DESVIO PADRÃO	1.01		

PADRONIZAÇÃO

É o processo de transformar uma variável de modo que ela fique com média zero e desvio padrão 1.

Se X é a variável que desejamos padronizar, então a VARIÁVEL PADRONIZADA será:

$$\frac{x - \bar{x}}{s_X}$$

Conceitos-Chave

ESTATÍSTICAS - MEDIDAS DE POSIÇÃO - MÉDIA - SOMATÓRIOS - SOMA DOS QUADRADOS - QUADRADO DA SOMA - MEDIANA - MODA - QUARTIS - PERCENTIS - MEDIDAS DE DISPERSÃO - AMPLITUDE DE VARIAÇÃO - DISTÂNCIA INTERQUARTIL - DESVIO - SOMA DE QUADRADOS - VARIÂNCIA - DESVIO PADRÃO - COEFICIENTE DE VARIAÇÃO - TRANSFORMAÇÃO LINEAR - PADRONIZAÇÃO

Leitura Essencial

[IEMMA, 1992] Cap. VI: p.87-103. Cap. VII: p109-119. Cap. VIII: p123-126.

Leituras Adicionais

[ROCHA, 1975] Cap. 9: p.69-89 (especialmente p87-88). Cap. 10: p90-94.

[AZEVEDO & CAMPOS, 1987] Cap. 6: p.127-172..

Exercícios

5.1 Considere os seguintes diâmetros (cm) das árvores de uma floresta de *Pinus caribaea hondurensis*.

21.8	12.3	35.5	32.2	22.3	24.9	18.2	13.2	20.3	10.7
23.9	10.8	19.3	21.4	16.4	18.2	27.3	26.4	9.4	19.4
27.7	16.0	12.7	33.0	15.1	13.2	10.8	18.0	14.5	20.1
21.8	11.5	16.3	17.5	28.8	17.0	21.7	28.0	9.0	17.6

a) Calcule a média. b) Calcule a mediana. c) Encontre o primeiro e terceiro quartil.
d) Encontre o percentil 75%. e) Calcule a variância f) Calcule o desvio padrão.
g) Calcule o CV.

5.2 Considere a produção de madeira (m^3/ha) de diferentes talhões de *Eucalyptus camaldulensis* aos 7 anos.

220	223	218	216	228	246	250	261	276	239
174	232	171	225	208	245	248	214	204	270
201	271	238	132	270	256	189	199	143	218
221	271	183	148	221	275	186	208	198	237
223	201	245	198	245	166	228	204	224	166

a) Calcule a média. b) Calcule a mediana. c) Encontre o primeiro e terceiro quartil.
d) Encontre o percentil 75%. e) Calcule a variância f) Calcule o desvio padrão.
g) Calcule o CV.

5.3 Considere os seguintes dados de resistência a compressão (kg/cm^2) de amostras de madeira de *Ocotea porosa* (peroba).

```
2765 2477 2755 3225 3387 2351 2160
1790 2802 1625 3062 1929 2129 2280
2594 1764 2761 3070 2609 1877 1259
2674 2123 3591 1264 1630 2347 1853
2170 2117 2433 2464 2813 1917 2541
2784 3159 2275 2260 2172 3034 2015
1758 1407 2384 2175  609 2572 3003
```

a) Calcule a média. b) Calcule a mediana. c) Encontre o primeiro e terceiro quartil.
d) Encontre o percentil 75%. e) Calcule a variância f) Calcule o desvio padrão.
g) Calcule o CV h) Encontre o percentil 20%.

5.4 Um fornecedor de peças para estruturas de madeira recebeu um lote de peroba de onde foram retiradas as amostras do exercício anterior. Para fins de venda das peças qual é a <u>resitência à compressão do lote inteiro</u> que este vendedor anunciar: média, mediana, percentil 75% ou percentil 20% ? Por que ?

5.5 Uma série de amostras de solo resultaram na seguinte distribuição para os valores de pH de um solo latossol vermelho-amarelo.

Classes de pH	Freqüência
4.6 - 5.0	7
5.0 - 5.4	14
5.4 - 5.8	23
5.8 - 6.2	32
6.2 - 6.6	19
6.6 - 7.0	8
7.0 - 7.2	2
Total	105

Encontre: a) a média b) a mediana c) o primeiro e terceiro quartil
d) os percentis: 10%, 35%, 75%, 95% e) a variância
f) o desvio padrão g) o coeficiente de variação.

5.6 A distribuição dos diâmetros das árvores de uma floresta nativa segue a seguinte tabela:

Classes de Diâmetro (cm)	Freqüência
10 - 20	351
20 - 30	160
30 - 40	86
40 - 50	40
50 - 60	20
60 - 70	4
70 - 80	4
80 - 90	3
90 - 100	1
Total	669

a) Qual o diâmetro médio e o diâmetro mediano desta floresta? Qual dos dois é maior?
b) Qual o desvio padrão dos diâmetros desta floresta ?

c) Um Engenheiro Florestal deseja fazer um corte seletivo retirando 20% das maiores árvores. Qual o diâmetro mínimo das árvores a serem removidas ?
d) Uma Engenheira Florestal deseja fazer um corte seletivo retirando 40% das menores árvores. Qual o diâmetro mínimo das árvores remanescentes?

5.7 A distribuição dos diâmetros das árvores de uma floresta plantada segue a seguinte tabela:

Classes de Diâmetro (cm)	Freqüência
4 - 10	10
10 - 14	172
14 - 18	210
18 - 22	144
22 - 26	125
26 - 30	67
Total	728

a) Qual o diâmetro médio e o diâmetro mediano desta floresta? Qual dos dois é maior?
b) Qual o desvio padrão dos diâmetros desta floresta ?
c) Um Engenheiro Florestal deseja fazer um corte seletivo retirando 20% das maiores árvores. Qual o diâmetro mínimo das árvores a serem removidas ?
d) Uma Engenheira Florestal deseja fazer um corte seletivo retirando 40% das menores árvores. Qual o diâmetro mínimo das árvores remanescentes?

5.8 Considere as seguintes quantias: $\sum_{i=1}^{n} x_i = 35;\ \sum_{i=1}^{n} y_i = 57;\ n = 23;\ k = 8.$

Encontre:

a) $x_1 + x_2 + x_3 + \ldots + x_n$ b) $y_1 + y_2 + y_3 + \ldots + y_n$ c) $\sum_{i=1}^{n} k$

d) $\sum_{i=1}^{n} kx_i$ e) $\sum_{i=1}^{n} x_i^2$ f) $\left[\sum_{i=1}^{n} x_i\right]^2$

g) $\sum_{i=1}^{n} x_i y_i$ h) $\sum_{i=1}^{n} (x_i - y_i)$ i) $\sum_{i=1}^{n} (x_i - y_i)^2$

j) $\sum_{i=1}^{n} (x_i - k)$ k) $\sum_{i=1}^{n} (x_i - k)^2$ l) $y_1^2 + y_2^2 + y_3^2 + \ldots + y_n^2$

5.9 Com base nas propriedades dos somatórios, deduza a fórmula prática para cálculo da Soma de Quadrados dos Desvios.

5.10 Com base nas propriedades dos somatórios, demonstre que a média da variável $Y = aX + b$ (transformação linear de X) é de fato $\bar{y} = a\bar{x} + b$, onde a e b são constantes e $\bar{x}$ é a média da variável X.

6. Gerando Informação: "Amostragem"

Introdução: Por que Amostrar ?

Quando se deseja informação sobre uma dada situação, o que vem à mente da maioria das pessoas é obter a informação de **toda** a população de interesse. Quando observamos ou medimos todos os indivíduos da população de interesse, realizamos um **CENSO**.

Mas será que o censo é sempre o melhor método para obtermos informações úteis?

TESTE:
Observe o parágrafo abaixo por 30 segundos e conte quantos "s" ocorrem.

> SEMPRE QUE SE FAZ AMOSTRAGEM
> DEVEMOS TER EM MENTE A
> REPRESENTATIVIDADE
> DAS OBSERVAÇÕES REALIZADAS.

Será que se chegou ao número certo? Tentemos novamente:

> sempre que se faz amostragem
> devemos ter em mente a
> representatividade
> das observações realizadas.

E agroa, se chegou ao número certo? Tentemos mais uma vez:

> **S**empre que **S**e faz amo**S**tragem
> devemo**S** ter em mente a
> repre**S**entatividade
> da**S** ob**S**ervaçõe**S** realizada**S**.

Certamente neste último exemplo se verifica rapidamente que existem 9 "S" na parágrafo.

Por que não foi fácil vê-los rapidamente no primeiro exemplo ?

Limitações do Censo

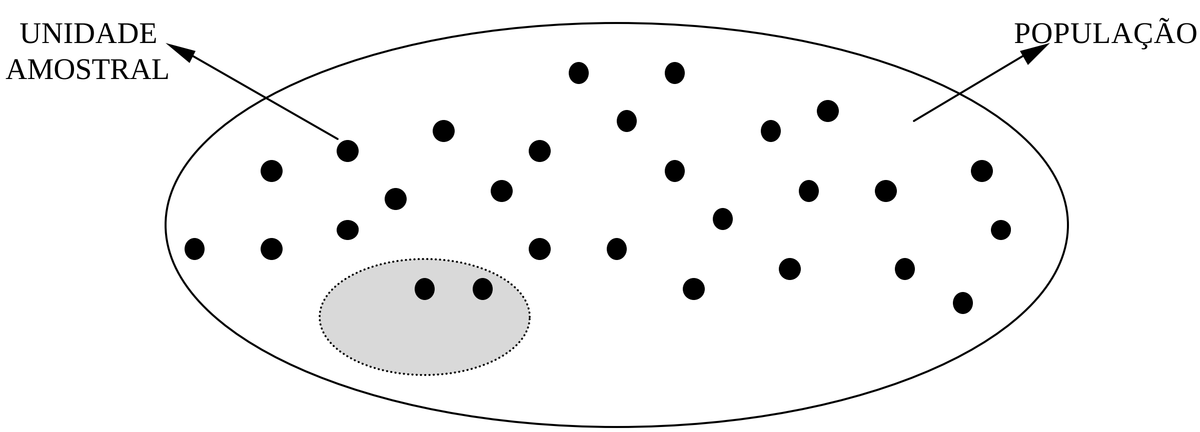
- Ineficaz: maior possibilidade de erro dada a quantidade de observações/medições.
- Impossível: no diagnóstico e monitoramento de populações extremante dinâmicas.
- Impraticável: custo e tempo de execução em geral são excessivos.

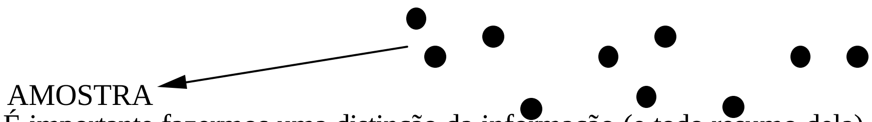
Terminologia em Amostragem.

> AMOSTRAGEM
>
> Consiste em **observar parte** da população para se fazer inferências (concluir algo) a respeito de **toda** a população.

Alguns termos utilizados em amostragem são semelhante a termos utilizados no dia-a-dia, mas o significado destes termos é bem preciso no caso de amostragem:

POPULAÇÃO:	Conjunto completo de ítens sobre os quais desejamos alguma informação.
UNIDADE AMOSTRAL:	Um ítem individual da população.
AMOSTRA:	Parte da população que é de fato utilizada para se obter a informação desejada.
VARIÁVEL:	A característica de interesse que é medida/observada em cada unidade amostral.





É importante fazermos uma distinção da informação (e todo resumo dela) que se obtem a partir da POPULAÇÃO e que se obtem a partir da AMOSTRA.

População → **PARÂMETRO** → Resumo numérico que se calcularia a partir da população e que descreve uma característica de interesse

Amostra → **ESTATÍSTICA** → Resumo numérico que é calculado a partir da amostra.

Exemplo 6.A: Aeróbica e Audição

> *"Médicos do Hospital Henry Ford (Detroid, EUA) estudaram 125 classes de aeróbica em 5 academias de ginástica e encontram que o nível de ruído em 60% das classes excedia os limites aceitáveis de ruído."*

População: ______________________________________

Unidade Amostral: ______________________________

Variável: ____

Amostra: ____

Parâmetro: ______________________________________

Estatística: ______________________________________

Exemplo 6.B: Ocorrência de Doenças numa Floresta Plantada

O objetivo de um levantamento florestal é determinar a taxa média de ocorrência de seca-do-ponteiro numa floresta de *Eucalyptus saligna*.

População: ______________________________________

Unidade Amostral: ______________________________

Variável: ____

Amostra: ____

Parâmetro: ______________________________________

Estatística: ______________________________________

Exemplo 6.C: Produção de Floresta Nativa

Para saber a produção (madeira para serraria) de uma Floresta Nativa de 5000*ha*. Assim, um Engenheiro Florestal mediu o volume de madeira (m^3/ha) em 52 parcelas de 1*ha*.

População: ______________________________________

Unidade Amostral: ______________________________

Variável: ____

Amostra: ____

Parâmetro: ______________________________________

Estatística: ______________________________________

Exemplo 6.D: Controle de Qualidade

Uma serraria vende lotes de 1000 tábuas madeira tropical nativa a um exportador. Antes do exportador enviar os lotes ao exterior ele verifica uma amostra de 20 tábuas por lote. A serraria e o exportador assinaram um contrato que determina que somente lotes com uma proporção igual ou inferior a 5% de peças defeituosas é aceitável.

População: ______________________________________

Unidade Amostral: ______________________________

Variável: ____

Amostra: ____

Parâmetro: ______________________________________

Estatística: ______________________________________

Exemplo 6.E: Estudo de Bioestatística

Suponhamos que a nossa população seja apenas um grupo de 5 alunos de Bioestatística e a variável de interesse seja o número de horas por semana dedicadas ao estudo da Bioestatítica (veja tabela ao lado).

- Calcule o tempo médio dedicado à Bioestatística nesta população (Parâmetro).

HORAS DE ESTUDO DE BIOESTATÍSTICA

Aluno	Horas de Estudo
1	1
2	3
3	4
4	2
5	5

Tomemos uma amostra de tamanho 2 desta população, qual o valor da Estatística se:

- Os alunos 1 e 4 forem selecionados?
- Os alunos 2 e 3 forem selecionados?
- Os alunos 1 e 5 forem selecionados?

O valor do Parâmetro e da estatística são sempre iguais?

População → **PARÂMETRO** → Valor **FIXO** e "**DESCONHECIDO**".

Amostra → **ESTATÍSTICA** → Valor **VARIÁVEL** (depende da amostra obtida) e **CONHECIDO** (para a amostra obtida).

Boa Amostragem? Viés e Precisão

A tendência natural das pessoas é buscar a informação do modo mais "***conveniente***".

→ Compra-se a caixa de morango com os morangos mais bonitos por cima. E os de baixo ?

→ A minha opinião sobre como anda a administração da cidade depende dos lugares onde vou e das ruas por onde passo. Será que os lugares e ruas que conheço representam todos os lugares/ruas da cidade?

→ A minha noção da opinião pública é baseada na opinião das pessoas com que convivo. Será que os meus conhecidos e amigos representam a população brasileira?

AMOSTRAGEM DE CONVENIÊNCIA

São observadas as unidades mais acessíveis da população . A seleção das unidades da amostra é realizada segundo a "conveniência".

As unidade "inconvenientes" de serem observadas jamais serão amostradas. → Não importa quão grande seja a amostra, amostragem de conveniência nunca será totalmente representativa da população. ⇒ Amostragem de Conveniência é sempre ENVIESADA, isto é, possui VIÉS.

VIÉS

Tendenciosidade presente no processo de amostragem que impede que a amostra seja perfeitamente representativa da população, independentemente do tamanho da amostra.

VIÉS DE SELEÇAO

Tendência sistemática de subrepresentar ou excluir da amostra um certo tipo de unidade.

Ex.: Pesquisas de opinião por telefone.

VIÉS DE NÃO-RESPOSTA

Numa pesquisa de opinião, o resultado de pessoas não responderem às questões do entrevistador.

Ex.: Pesquisas de opinião pelo correio.

VIÉS DE RESPOSTA

Numa pesquisa de opinião, a forma de apresentação das questões são e a atitude do entrevistador podem influenciar a resposta.

Ex.: Ordem do nome dos candidados numa pesquisa de intensão de voto.

Exemplo 6.F: Você decide!

Um programa de televisão conduz a seguinte pesquisa de opinião:

> *Pesquisa científica mostra que o uso do cinto de segurança na cidade reduz o número de acidentes com acidentados graves. Se você for a favor do uso do cinto de segurança na cidade ligue para 0-800-552-4949, se for contra ligue para 0-800-552-4950, ligação gratuita.*

- Enumere as razões pelas quais esta pesquisa de opinião é certamente enviesada.

Exemplo 6.G: Ipê Florido.

Uma Engenheira Florestal deseja estudar o padrão de floração de ipê amarelo numa mata nativa. Para isso, ela selecionou aleatoriamente algumas árvores de ipê amarelo na mata e visitou todas as árvores num mesmo dia, classificando cada árvore numa escala qualitativa de intensidade de florescimento.

- Existe algum tipo de viés neste estudo?

Exemplo 6.H: Dez Amostras.

Estudando a influência do sistema de amostragem no estrutura de tamanho das árvores de uma floresta tropical, um Ecologista tomou 10 Amostras compostas de 5 árvores, 10 amostras compostas de 20 árvores e 10 amostras compostas de 40 árvores (tabela ao lado).

- Qual o efeito do tamanho da amostra (número de árvores) no resultado nos resultados obtidos?

MÉDIA DO DIÂMETRO DE ÁRVORES AMOSTRADAS NUMA FLORESTA NATIVA

Amostras	Tamanho das Amostras (número de árvores) 5	20	40
1	26.6	34.1	30.5
2	32.9	31.2	29.8
3	39.2	40.1	37.3
4	30.3	30.4	29.4
5	31.8	29.7	32.7
6	48.0	38.8	31.5
7	37.7	37.3	29.3
8	20.0	35.6	30.6
9	31.8	37.7	31.2
10	33.2	26.7	29.3

Amostras de maior tamanho sempre produzem Estatísticas com menor variabilidade entre amostras.

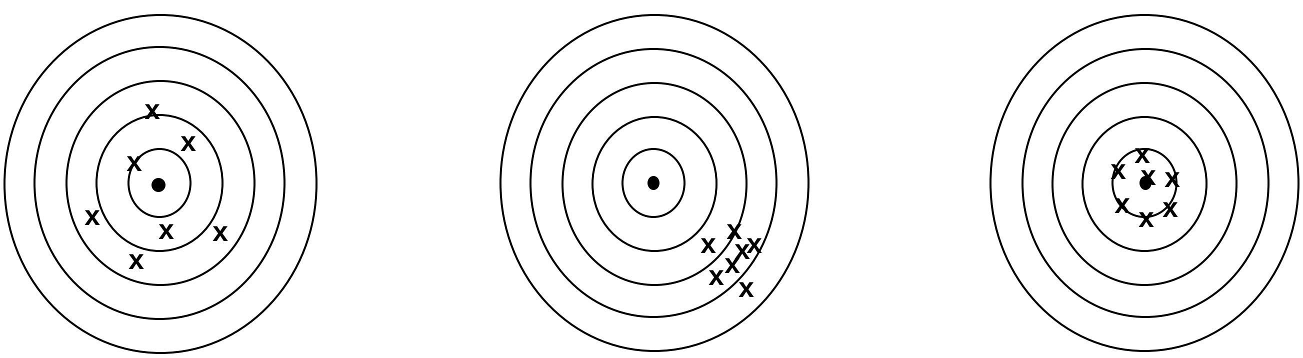
Amostras maiores são sempre **mais precisas**.

Para um mesmo tamanho de amostra, sistemas diferentes de amostragem podem resultar em precisão diferente.

PRECISÃO

Capacidade dos sistemas de amostragem de gerar Estatísticas com pequena variabilidade entre amostras (de mesmo tamanho) tomadas numa mesma população.

QUAL O SISTEMA DE AMOSTRAGEM IDEAL?



SEM VIÉS SEM PRECISÃO	COM VIÉS COM PRECISÃO	SEM VIÉS COM PRECISÃO

Exemplo 6.I: Fragmentação Florestal.

Uma equipe de cientístas deseja monitorar as mudanças ecológicas numa floresta após a fragmentação desta devido a conversão da área floresta para área agrícola. A equipe se encontra dividida entre duas alternativas de amostragem:

- Monitorar um grande número de fragmentos que sejam de fácil acesso e que representem o atual estado da região.
- Monitorar um número pequeno de fragmentos que sejam uma amostra representativa da floresta original, independentemente da facilidade de acesso.

Quais as vantagens e desvantagens das duas alternativas?

Exemplo 6.J: Instrumento de Medição.

Um Engenheiro Florestal dispõem de dois instrumentos para medir a altura das árvores. O instrumento A tem alta precisão, mas apresenta um viés sistemático que independe da altura da árvore sendo medida. O instrumento B tem precisão inferior ao instrumento A, mas não apresenta qualquer tipo de viés.
Este Engenheiro foi chamado para realizar medições em duas situações:

- Experimento que compara diferentes híbridos de *Pinus caribaea* - *Pinus elliottii*.
- Levantamento do volume de madeira em pé numa floresta a ser vendida.

Qual dos dois instrumentos deve ser utilizado em cada situação? Por que?

Métodos de Amostragem

Existe vários Métodos de Amostragem utilizados nas Ciências Florestais. Todos têm como objetivo gerar informações sem viés e com a maior precisão possível.

Métodos de Amostragem utilizados nas Ciências Florestais.

São métodos objetivos (sem interferência das pessoas que os utilizam) e imparciais.

Todas unidades da população possuem uma certa chance de serem selecionadas na amostra.

Esta chance é conhecida e é diferente de zero.

Para estes métodos seremos capazes de calcular a precisão de uma dada estatística para representar o parâmetro da população.

MÉTODOS DE AMOSTRAGEM:

- **Amostragem Aleatória Simples**
- **Amostragem Aleatória Estratificada**
- **Amostragem Sistemática**
- **Amostragem em Múltiplos Estágios**

Amostragem Aleatória Simples (AAS)

AAS

Uma amostra de um dado tamanho tem a mesma chance de ser selecionada que qualquer outra amostra possível de mesmo tamanho.

Amostras de diferentes tamanhos podem ter chances diferentes.

Exemplo Clássico:
Bolas de mesmo tamanho, peso, superfície, etc., numa urna. Qualquer amostra de 3 bolas tem exatamente a mesma chance de ser obtida.

O **PROCESSO ALEATÓRIO** de seleção das unidades elimina qualquer interferência de escolha humana na escolha dos elementos da amostra.

UM PROCESSO ALEATÓRIO ADEQUADO
É A GARANTIA DE AUSÊNCIA DE VIÉS.

Exemplo de Processo Aleatório: Bingo.

Uma urna contendo os números: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
Antes de cada sorteio, os números são "mixturados" na urna.
Ao sortear um número ele pode:
* ficar em separado ==> **Amostragem Sem Reposição**;
* retornar à urna ==> **Amostragem Com Reposição**.

Exemplo de Processo Aleatório: Tabela de Números Aleatórios.

Uma Tabela de Números Aleatórios pode ser entendida como o bingo acima, usando a amostragem com reposição.

TABELA DE NÚMEROS ALEATÓRIOS
(Anexo A)

É uma lista de 10 dígitos (0,1,2,3,4,5,6,7,8,9) que possue as seguintes propriedades:

(1) O dígito em qualquer posição na lista tem a mesma chance de ser um dos dígitos: 0,1,2,3,4,5,6,7,8 ou 9.

(2) Os dígitos em diferentes posições na lista são independentes, no sentido de que o valor de um dígito não influencia o valor de outro.

Para facilitar o seu uso, a Tabela de números aleatórios é organizada em linhas e colunas.

As propriedades da Tabela de Números Aleatório implicam que:

- Qualquer par de dígitos tem a mesma chance de ser um dos 100 possíveis pares: 00, 01, **. . .** , 99.
- Qualquer trio de dígitos tem a mesma chance de ser um dos 1000 possíveis trios: 000, 001, **. . .** , 999.
- Qualquer **. . .**

Como utilizar Números Aleatórios numa Amostragem?

TERMINOLOGIA

N (maiúsculo) denota o tamanho da população.

n (minúsculo) denota o tamanho da amostra.

A) Rotular todas N unidades da população:

1. Designar cada unidade na população por um rótulo composto de dígitos.

2. Sempre utilizar o menor número possível de dígitos nos rótulos:
   Para 10 ou menos unidades, utilizar rótulos de um dígito: 0, 1, . . . , 9.
   Para mais de 10 e menos de 100, utilizar dois dígitos nos rótulos: 00, 01, . . . , 99.

3. Todos os rótulos devem possuir o mesmo número de dígitos.

4. Cada unidade deve ser designada pelo mesmo número de rótulos.

<u>B) Selecionar uma amostra de tamanho n</u>:

1. Iniciar num ponto aleatório da Tabela de Números Aleatórios.

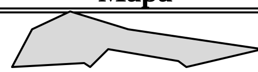
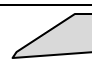
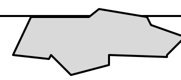
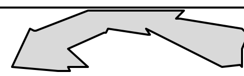
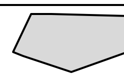
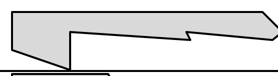
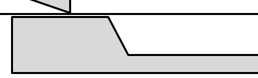
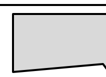
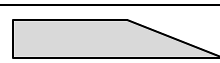
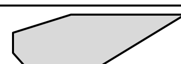
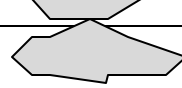
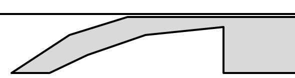
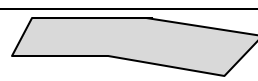
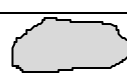
2. Ler os rótulos sistematicamente seguindo as linhas ou colunas da Tabela.

3. Se um rótulo já lido surgir novamente ou se aparecer um rótulo que não está presente na população, ignorá-lo e seguir para o próximo.

4. Continuar até que n unidades sejam selecionadas.

Exemplo 6.K : Amostrando Fragmentos

Realizar uma amostra aleatória simples de tamanho 4 dos fragmentos:

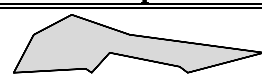
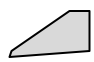
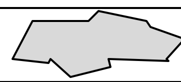
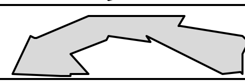
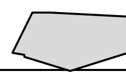
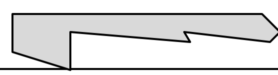
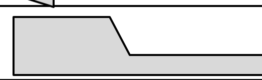
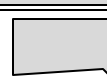
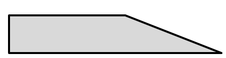
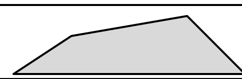
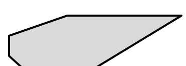
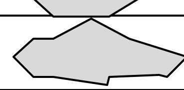
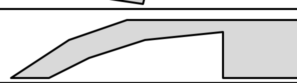
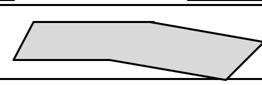
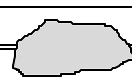
Fragmento	Rótulo	Mapa	Área (ha)
Fazenda Santa Rita	01		35
Fazenda Santa Rita 2	02		2
Piraquera	03		17
Fazenda Sertão	04		10
Fazenda Água Clara	05		19
Fazenda Tijuco Preto	06		4
Fazenda Dona Zilma	07		40
Sítio Arueira	08		1
Fazenda Água Funda	09		7
Sítio Santa Clara	10		9
Sítio Santo Antônio	11		12
Fazenda Conceição	12		10
Sítio do Bento	13		18
Fazenda Pau Queimado	14		34
Fazenda Gibeira	15		21

Quando o tamanho da população é muito pequeno em relação ao número de rótulos, muito tempo é gasto saltando rótulos que não são úteis. Para evitar isto podemos utilizar o seguinte procedimento:

<u>Selecionar uma amostra de tamanho n quando o número de rótulos é muito maior que o tamanho da população (N)</u>:
1. Iniciar num ponto aleatório da Tabela de Números Aleatórios.
2. Ler os rótulos sistematicamente seguindo as linhas ou colunas da Tabela.
3. Se um rótulo já lido surgir, ignorá-lo e seguir para o próximo.
4. Se aparecer um rótulo maior que o maior rótulo utilizado**, dividir o rótulo que surgiu por N e tomar o resto da divisão como rótulo**.
5. Continuar até que n unidades sejam selecionadas.

Exemplo 6.L : Amostrando Fragmentos II

Realizar uma amostra aleatória simples de tamanho 4 **<u>com base na área dos fragmentos</u>**:

Fragmento	Rótulo	Mapa	Área (ha)	Área Acumulada
Fazenda Santa Rita	01		35	35
Fazenda Santa Rita 2	02		2	37
Piraquera	03		17	54
Fazenda Sertão	04		10	64
Fazenda Água Clara	05		19	83
Fazenda Tijuco Preto	06		4	87
Fazenda Dona Zilma	07		40	127
Sítio Arueira	08		1	128
Fazenda Água Funda	09		7	135
Sítio Santa Clara	10		9	144
Sítio Santo Antônio	11		12	156
Fazenda Conceição	12		10	166
Sítio do Bento	13		18	184
Fazenda Pau Queimado	14		34	218
Fazenda	15		21	239

Gibeira

Exemplo 6.M : Proporção de Mulheres no Curso de Bioestatística

A população é definida como os alunos em sala de aula, e a questão é a a proporção de mulheres na população.

- Números de mulheres na população: contagem = ____ .
- Número de alunos: N = ____ .
- Proporção na população (parâmetro): p = contagem / N = ___________________ .

Tome uma AAS de tamanho 4:

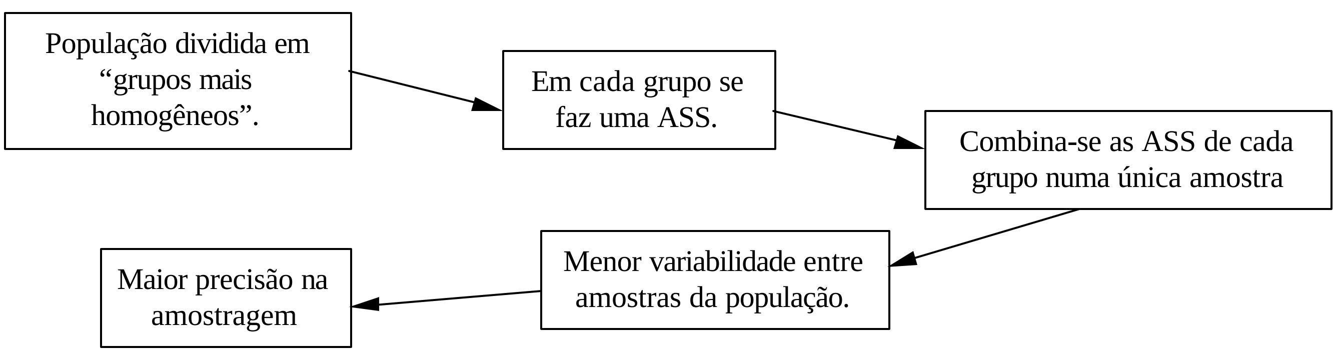
- Números de mulheres na amostra: contagem = ____ .
- Número de alunos: n = ____ .
- Proporção na população (estatística): $\hat{p}$ = contagem / n = ____________________ .

Parâmetro e estatística são iguais (p = $\hat{p}$) ?

Exemplo 6.N : Amostrando Fragmentos III

Realizar uma ASS de tamanho 4, tomando cada fragmento como uma unidade.
Com base nessa amostra estime: a área média dos fragmentos, a área mediana dos fragmentos.

Amostragem Aleatória Estratificada (AAE)



AMOSTRAGEM ALEATÓRIA ESTRATIFICADA (AAE)

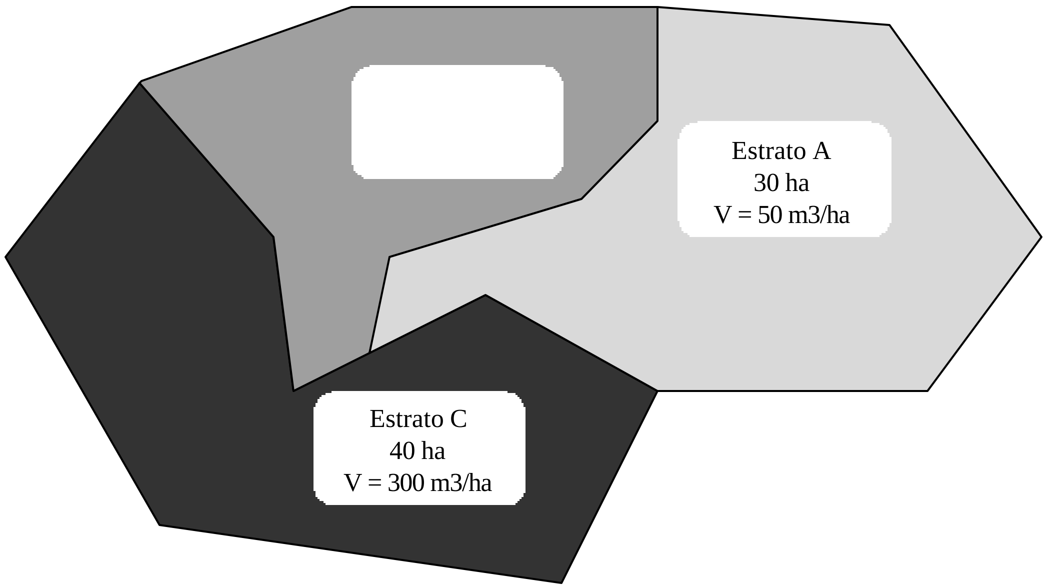
A população é sub-divida em grupos mutuamente exclusivos chamados **ESTRATOS**. Uma ASS é tomada dentro de cada estrato, e as estimativas dos estrato são combinadas.

Como os grupos são mutuamente exclusivos, cada unidade da população pertence a um único estrato.

Se os estratos forem internamente mais homogêneos que a população em termos da variável sendo medida, a AAE será mais precisa que a ASS.

Exemplo 6.O: Produção em Floresta Plantada.

Suponhamos que uma floresta plantada se comporte conforme a figura abaixo. Façamos também a pressuposição irreal de que os estratos são “perfeitamente homogêneos”, isto é, cada área de 1ha dentro de cada estrato tem a mesma produção (conforme a figura).



A) Realize 2 AAS de tamanho 6 com base na área total da floresta (100ha).
Qual a variância das estimativas da produção média (m3/ha)?

B) Realize 2 AAE de tamanho 6 (duas unidades por estrato).
Como a média da AAE é calculada?

$$\overline{x} = \left(\frac{30}{100}\right)\overline{x}_A + \left(\frac{30}{100}\right)\overline{x}_B + \left(\frac{40}{100}\right)\overline{x}_C =$$

Qual a variância das estimativas da produção média (m3/ha)?

Am

CÁCULO DA MÉDIA NA AAE

$$\overline{x} = \left(\frac{N_1}{N}\right)\overline{x}_1 + \left(\frac{N_2}{N}\right)\overline{x}_2 + \ldots + \left(\frac{N_k}{N}\right)\overline{x}_k = \left(\frac{1}{N}\right)\sum_{i=1}^{k} N_i \overline{x}_i$$

$\overline{x}$ é a média da AAE

$\overline{x}_i$ é a média estimada para o estrato *i*.

N é o tamanho da população (número total de unidades)

N_i é o tamanho do estrato *i* (número de unidades no estrato k)

A amostragem sistemática é uma maneira mais prática de selecionar as unidades de uma população.

AMOSTRAGEM SISTEMÁTICA

Inicialmente, as unidades da população são colocadas numa certa ordem..

Para uma amostra sistemática de "**1-em-cada-k-unidades**", você primeiro seleciona aleatoriamente uma das k primeiras unidades. A partir da unidade selecionada toma-se sempre a k-ésima unidade.

NOTE: somente a primeira unidade é selecionada aleatoriamente, todas as demais são tomadas sistematicamente.

Exemplo 6.P: Amostra Sistemática de 1-em-4.

Suponha que a população sendo estudada possui 19 unidades, tendo com rótulos as letras A até S.
Selecione uma amostra sistemática de 1-em-4.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19

Qual a chance da letra A ser selecionada? ______________________________
Qual a chance da letra G ser selecionada ? ______________________________
Qual a chance da letra N ser selecionada ? ______________________________

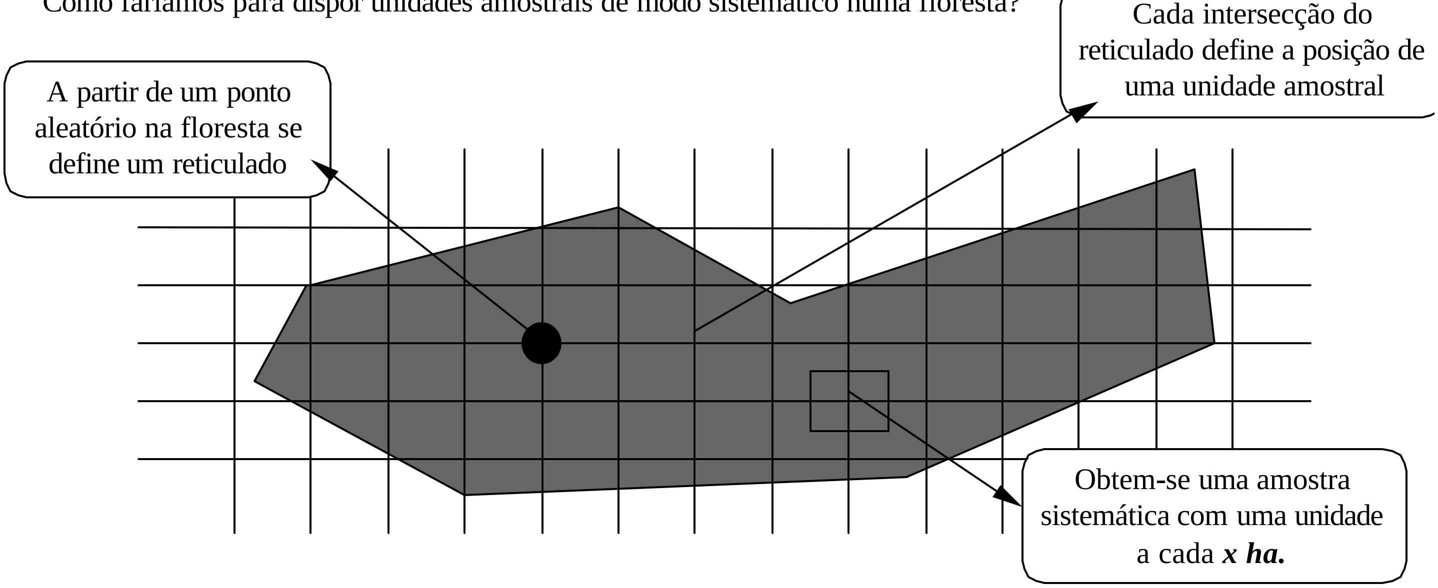
Uma amostra sistemática é uma amostra aleatória simples? Explique.

Uma amostra sistemática é uma amostra estratificada? Explique.

O tamanho da amostra é fixo numa amostra sistemática? Explique.

Exemplo 6.Q: Amostra Sistemática numa Floresta.

Como faríamos para dispor unidades amostrais de modo sistemático numa floresta?



Amostragem em Múltiplos Estágios (AME)

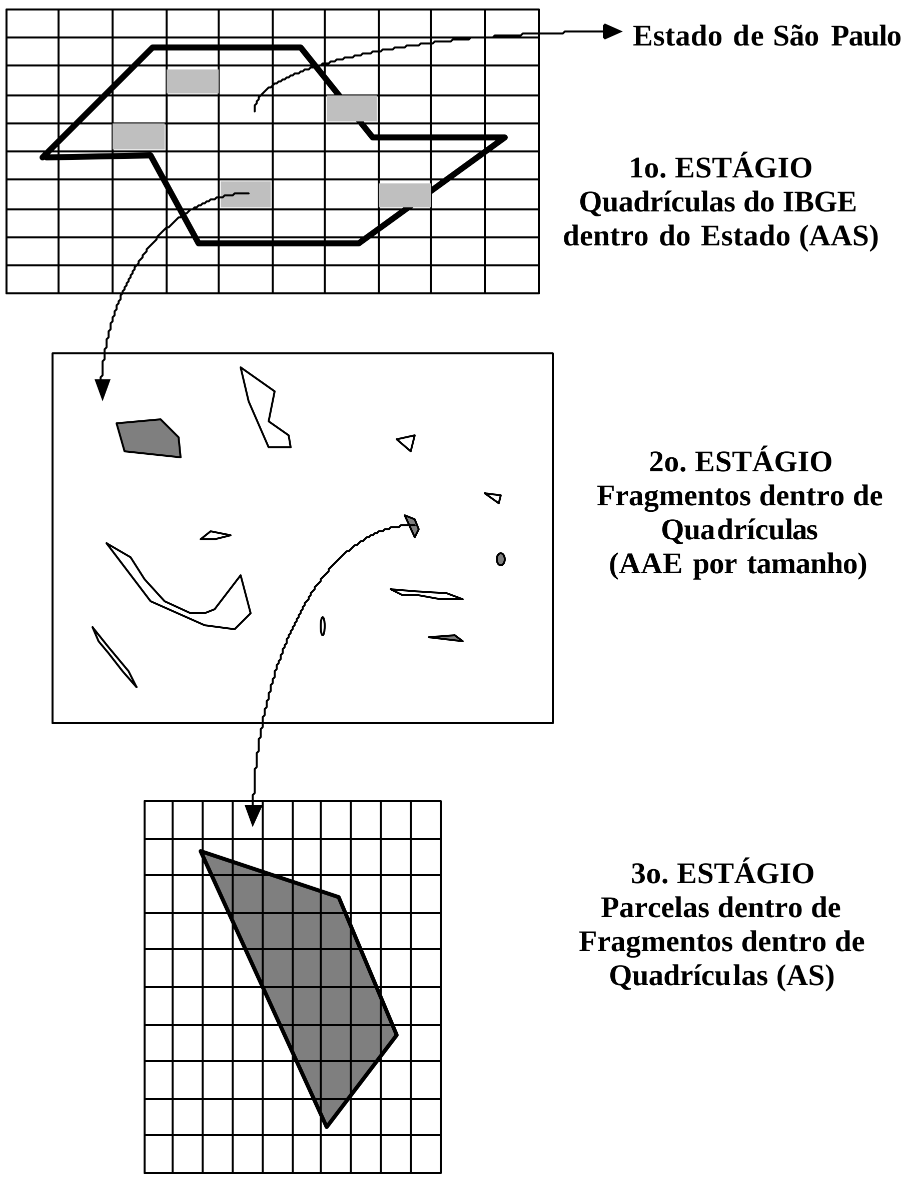
Como o nome sugere, na amostragem em múltiplos estágios o processo de amostragem aleatória é repetido em vários estágios hierarquizados.

Em cada estágio, qualquer um dos métodos de amostragem discutidos pode ser utilizado.

Exemplo 6.R: Levantamento Regional

O DPRN (Departamento de Proteção dos Recursos Naturais) do Estado de São Paulo deseja realizar um levantamento rápido para diagnosticar a situação dos fragmentos florestais em todo Estado.

Qual a melhor forma de amostrar?



Conceitos-Chave

CENSU - AMOSTRAGEM - POPULAÇÃO - UNIDADE AMOSTRAL - AMOSTRA - VARIÁVEL - PARÂMETRO - ESTATÍSTICA - AMOSTRAGEM DE CONVENIÊNCIA - VIÉS - VIÉS DE SELEÇÃO - VIÉS DE NÃO-RESPOSTA - VIÉS DE RESPOSTA - PRECISÃO - MÉTODOS DE AMOSTRAGEM - AMOSTRAGEM ALEATÓRIA SIMPLES - PROCESSO ALEATÓRIO - TABELA DE NÚMEROS ALEATÓRIOS - AMOSTRAGEM ALEATÓRIA ESTRATIFICADA - AMOSTRAGEM SISTEMÁTICA - AMOSTRAGEM EM MÚLTIPLOS ESTÁGIOS

Leitura Essencial

[OLIVEIRA, 1977] p.39-43.

Exercícios

6.1 O Diretor deseja saber a opinião do corpo discente sobre uma nova proposta de procedimento de matrícula. Para obter a informação rapidamente decidiu-se por amostragem.
a) Qual é a população? b) Quais são as unidades amostrais? c) Qual é a variável? d) Qual método de amostragem deve ser utilizado? Explique.

6.2 Uma empresa florestal pretende definir o nível de infestação de formigas nas suas florestas.
a) Qual é a população? b) Quais são as unidades amostrais? c) Qual é a variável? d) Qual método de amostragem deve ser utilizado? Explique.

6.3 Um candidato a governador do Estado deseja saber suas chances de vencer na próxima eleição e encomendou uma pesquisa de intensão de voto.
a) Qual é a população? b) Quais são as unidades amostrais? c) Qual é a variável? d) Qual método de amostragem deve ser utilizado? Explique.

6.4 Uma Engenheira Florestal realizou um estudo sobre o vandalismo nas trilhas de um Parque Nacional. Nesse estudo ela se aproximava anonimamente a grupos de pessoas caminhando nos fins-de-semana e fazia a caminhada com o grupo observando-os. Existe algum tipo de viés neste estudo? Qual? Explique.

6.5 Num estudo de ciclagem de nutrientes, o pequisador deseja saber o teor médio de fósforo na serrapilheira de uma floresta. Para isso ele seleciona pequenas áreas de $1m^2$ distantes 3 metros da trilha utilizada adentrar na floresta. Em cada área de $1m^2$, toda serrapilheira é removida e levada ao laboratório para análise. Existe algum tipo de viés neste estudo? Qual? Explique.

6.6 Num levantamento da avifauna do sub-bosque de uma mata seca no Estado de Mato Grosso, um pesquisador utiliza um gravador de alta sensibilidade para gravar o canto dos pássaros. Posteriormente, as fitas são analisadas em laboratório para identificação das espécies. Existe algum tipo de viés neste levantamento? Qual? Explique.

6.7 A Empresa Abril Cultura e Revista Playboy realiza anualmente o ranking das melhores escolas de Engenharia Florestal. O ranking se baseia num formulário enviado a uma amostra aleatória de profissionais formados que são convidados a classificar a escola onde se formaram. Existe algum tipo de viés neste levantamento? Qual? Explique.

6.8 A tabela abaixo mostra a produção por parcelas de 1ha de uma floresta de *Eucalyptus grandis*.

a) Realize 5 AAS de tamanho 8. Em cada amostra estime a produção média da floresta.
b) Realize 5 AAE de tamanho 8 (2 unidades/estrato). Em cada amostra estime a produção média da floresta.
c) Compare a variação das estimativas obtidas pela AAS e pela AAE, em torno do valor do parâmetro. Qual dos métodos apresenta maior precisão. Explique.

Idade (anos)	**Parcela (1ha)**	**Produção (m3/ha)**
3	1	50
	2	55
	3	56
	4	67
	5	48
	6	69
4	1	150
	2	148
	3	120
	4	160
	5	130
5	1	250
	2	210
	3	205
6	1	310
	2	280
	3	300
	4	320
	5	340

6.9 Um Engenheiro Florestal deseja realizar uma AAS de tamanho 3 no talhão da figura abaixo. Cada unidade amostral consiste de uma parcela de 20 x 30 m ($600m^2$). Realize essa amostragem. Explique detalhadamente o processo utilizado.

ESTRADA

1000m

6.10 A base na tabela abaixo apresenta o diâmetro (cm) de uma população de 200 árvores.

a) Calcule o diâmetro médio da população.
b) Realize 5 amostras aleatórias simples de tamanho 10. Para cada amostra estime a média da população. Analise a distribuição das médias e sua relação com a média da população.
c) Realize 5 amostras sistemáticas de tamanho 10. Para cada amostra estime a média da população. Analise a distribuição das médias e sua relação com a média da população.
d) Existe alguma diferença entre a distribuição das estimativas obtidas por AAS e por AS? Explique.

Linhas	**Colunas**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	16.0	54.7	27.7	29.2	10.7	38.9	24.3	38.2	11.8	24.0
2	83.7	10.8	17.6	31.6	27.9	20.1	36.8	39.8	11.6	22.6
3	11.9	49.8	15.5	22.5	17.6	36.7	56.1	15.8	28.8	20.9
4	20.3	20.6	48.3	26.0	76.4	56.0	18.5	44.8	37.3	51.7
5	38.9	21.7	12.6	52.9	24.6	39.7	12.1	34.4	70.4	42.8
6	22.4	21.2	14.9	20.0	12.2	27.7	28.2	43.3	40.7	43.4
7	21.9	30.2	14.6	37.0	24.3	10.5	15.1	18.0	16.2	16.1
8	35.7	41.3	13.0	11.4	16.6	31.2	14.3	38.2	11.2	30.3
9	56.5	21.6	17.1	27.3	13.5	13.6	37.3	18.3	39.4	22.0
10	12.3	23.1	34.2	13.6	34.3	91.7	81.1	39.6	15.6	54.5
11	26.5	117.8	14.7	15.4	25.9	12.9	123.2	16.8	20.1	30.1
12	58.3	34.3	15.6	11.0	24.8	11.6	22.9	12.7	19.2	42.2
13	14.7	10.3	15.3	27.0	18.2	59.1	129.0	16.7	12.9	23.9
14	15.5	19.8	14.6	50.0	10.2	24.6	20.0	10.9	16.9	11.0
15	25.8	21.7	62.2	94.5	116.7	23.5	96.7	10.4	68.7	15.8
16	33.4	75.4	20.2	11.8	16.7	49.4	29.2	26.7	27.5	32.1
17	19.6	92.4	12.7	20.6	14.9	10.7	10.1	15.6	14.2	26.2
18	15.3	20.8	14.3	30.3	25.8	17.5	10.2	25.8	12.3	61.3
19	25.7	16.9	15.2	10.7	10.6	19.1	24.8	12.8	21.5	16.5
20	25.3	10.3	17.4	38.5	16.4	48.8	16.8	30.1	17.1	42.4

7.Informação e Incerteza: "Probabilidade"

Exemplo 7.A: Planejamento Familiar

Um casal planeja ter filhos até terem uma menina ou até terem quatro filhos.
Qual a probabilidade deste casal ter um a filha ?

Encontrando a probabilidade por simulação:

1. Utilize a tabela de números aleatórios e selecione 20 números:

```
1  2  7  0  3  7  1  4  5  6  3  5  0  6  2  7  2  1  5  9
```

2. Os números pares represetam meninas (F) e os ímpares meninos (M):

```
1  2  7  0  3  7  1  4  5  6  3  5  0  6  2  7  2  1  5  9
M  F  M  F  M  M  M  F  M  F  M  M  F  F  F  M  F  M  M  M
```

3. Agrupe os números de acordo com planejamento do casal para simular as diferentes possibilidades de filhos. Numere cada simulação como uma família:

```
1  2  7  0  3  7  1  4  5  6  3  5  0  6  2  7  2  1  5  9
M  F  M  F  M  M  M  F  M  F  M  M  F  F  F  M  F  M  M  M
1  1  2  2  3  3  3  3  4  4  5  5  5  6  7  8  8  9  9  9
```

No exemplo acima, a simulação gerou 9 famílias.

4. Calcule a probabilidade do casal ter uma filha: 9/9 = 1.

- Será que a probabilidade é de 100% ?
- Será que se selecionarmos um outro conjunto de 20 números aleatórios obteremos a mesma probabilidade***?***

Repetindo a simulação com vários grupos:

Grupo	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Total**
No. de Famílias																					
No. de Famílias c/ Filhas																					

Probabilidade por simulação = Total de Famílias / Total de Famílias c/ Filhas = _______/_______ = ____________

Probabilidade exata = 0.938

Exemplo 7.B: Loteria Federal

Num jogo semelhante à Loteria Federal um jogador comprou o bilhete 123456. Este jogador receberá:

R$ 50000,00	se o bilhete sorteado for	**123456**
R$ 5000,00	se o bilhete sorteado for	**X23456**
R$ 500,00	se o bilhete sorteado for	**XX3456**
R$ 50,00	se o bilhete sorteado for	**XXX456**

(X representa qualquer dígito sorteado não correspondente ao bilhete).

1. Elabore um sistema para simular uma rodada da loteria utilizando a tabela de números aleatórios.

2. Simule 5 rodadas da loteria e anote os números sorteados.

3. Em quantas rodadas o bilhete 123456 foi sorteado? Qual o valor dos prêmios?

4. Junte o resultado de suas simulações com o de outros alunos da sala:

Aluno	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**Total**
No. de rodadas																					
No. de rodadas c/ o bilhete sorteado																					

5. Com base nas simulações de todos os alunos, qual a probabilidade do bilhete 123456 ser sorteado?

Exemplo 7.C: Loteria Federal - II

Será que podemos calcular a probabilidade exata de se ganhar nesta loteria?

1. Quantos bilhetes são possíveis serem sorteados:

Ordem do Dígito	1o.	2o.	3o.	4o.	5o.	6o.	
Número de Dígitos Possíveis	10	10	10	10	10	10	$10^6 = 1000000$ Bilhetes

2. Quantos destes bilhetes pagarão algum prêmio?

R$ 50000,00	$\Rightarrow$	1
R$ 5000,00	$\Rightarrow$	9
R$ 500,00	$\Rightarrow$	9 x 9 = 81
R$ 50,00	$\Rightarrow$	9 x 9 x 9 = 729

$$\frac{1+9+81+729}{1000000} = \frac{820}{1000000} = 0.00082$$

Conceitos Básicos em Probabilidade

Experimento Aleatório

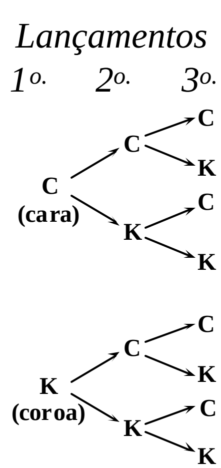
Processo estocástico (aleatório) em que o resultado final é incerto devido ao próprio processo.

- ◊ Número resultante do lançamento de um dado.
- ◊ Os números numa tabela de números aleatórios.
- ◊ O número de árvores com cancro numa parcela de floresta de eucalipto.
- ◊ As produtividades de diferentes áreas de floresta de Pinus que receberam adubação..
- ◊ As espécies de aves que são capturadas numa rede-neblina armada no sub-bosque de uma floresta nativa

Espaço Amostral (S)

Conjunto de **todos os resultados possíveis** de um experimento aleatório.
Alguns experimentos aleatórios podem ter um número ***infinito*** de resultados, por hora veremos experimentos apenas com um número ***contável*** de resultados.

◊ No lançamento de 3 moedas justas, o espaço amostral pode ser representado por:



S =

CCC
CCK
CKC
CKK
KCC
KCK
KKC
KKK

S = { CCC, CCK, CKC, CKK, KCC, KCK, KKC, KKK }

◊ Se no lançamento de 3 moedas justas, definirmos o resultado como o número de caras, o espaço amostral fica:

$$S = \{ 0, 1, 2, 3 \}$$

Dois aspectos fundamentais na representação do ESPAÇO AMOSTRAL

O espaço amostral não precisa ser necessariamente um conjunto de números, mas um **sistema de códigos** deve ser estabelecido quando os resultados não são numéricos.

A definição do que representa um **resultado** do experimento aleatório é fundamental na definição do espaço amostral.

Exemplo 7.D: Representando Espaços Amostrais

Lançamento de uma moeda:

$S = \{ \qquad\qquad\qquad \}$

Lançamento de dois dados:

$S = \{$ (,) (,) (,) (,) (,) (,)
(,) (,) (,) (,) (,) (,)
(,) (,) (,) (,) (,) (,)
(,) (,) (,) (,) (,) (,)
(,) (,) (,) (,) (,) (,)
(,) (,) (,) (,) (,) (,) $\}$

No lançamento de dois dados o resultado é definido como a soma dos números obtidos:

$S = \{ \qquad\qquad\qquad \}$

Um conjunto de 10 árvores, selecionadas aleatóriamente, é tomado numa floresta de *Eucalyptus urophylla*. O resultado é definido como o número de árvores bifurcadas.

$S = \{ \qquad\qquad\qquad \}$

Um estudante de Bioestatística Florestal é selecionado aleatoriamente. O resultado é definido como o número de horas que o estudante dedicou à Bioestatística nas últimas 24 horas.

$S = \{ \qquad\qquad\qquad \}$

Uma parcela de 1ha é locada aleatoriamente numa floresta nativa. O resultado é definido como o número de árvores na parcela.

$S = \{ \qquad\qquad\qquad \}$

Um aluno da ESALQ é selecionado ao acaso. O resultado é definido como o número de créditos que este aluno está cursando no atual semestre.

$S = \{ \qquad\qquad\qquad \}$

Um aluno da ESALQ é selecionado ao acaso. O resultado é definido como o número de créditos necessários para este aluno se graduar.

$S = \{ \qquad \qquad \qquad \qquad \qquad \qquad \}$

Evento

EVENTO

Eventos são **sub-conjuntos do espaço amostral** e são geralmente representados por letras maiúsculas, como A, B, C, ...

Experimento Aleatório:
Um dado é lançado.

Espaço Amostral:
$S = \{ 1, 2, 3, 4, 5, 6\}$

⇒

Eventos possíveis:

O resultado é um número impar:
$A = \{ 1, 3, 5 \}$

O resultado é menor do que 5:
$B = \{ 1, 2, 3, 4 \}$

O resultado é um número primo:
$C = \{ 1, 2, 3, 5 \}$

Se um dado é lançado e o resultado é um elemento de A, dizemos que o **evento A ocorre**.
Se o resultado não for um elemento de A, dizemos que o **evento A não ocorre**.

Conceito de Probabilidade

O termos "probabilidade" é utilizado com diferentes sentidos na linguagem do dia-a-dia:

⇒ A previsão do tempo indica 70% de probabilidade de chuvas no próximo fim-de-semana.

⇒ O jogador sente que esta noite as chances de levar uma "grande bolada" na roleta são altas.

⇒ Ao jogar uma moeda justa, a chance de coroa é de 1/2.

⇒ Se uma pessoa sofreu um acidente fatal, as chances de que este acidente ocorreu em casa são de 10.5%.

⇒ das pessoas consideram o elogio do chefe como um importante componente na satisfação profissional.

⇒ O sexismo latino-americano está comprovado: o nível de analfabetismo na América Latina é de 17% entre mulheres e de 14% entre homens.

⇒ Apesar das mulheres comporem aproximadamente 50% da poplação eleitoral na Europa, apenas 11% dos cargos eletivos eram ocupados por mulheres em 1995.

CONCEITO SUBJETIVO

As pessoas expressam julgamento subjetivo de possibilidades na forma percentual ou "probabilidade".

Embora apresentado de forma quantitativa, o conceito subjetivo não deve ser confundido com o conceito de probabilidade utilizado na Bioestatística.

Alguns métodos de gerenciamento, incorporam o conceito subjetivo de probabilidade sem, entretanto, eliminar o seu aspecto subjetivo.

CONCEITO CLÁSSICO OU "A PRIORI"

Pelo conceito clássico a probabilidade é definida com base em dados do experimento aleatório.

A probabilidade é obtida antes que do experimento ser realizado e, portanto, o nome "a priori".

O conceito clássico surgiu no século XVII a partir dos jogos de azar e define a probabilidade do evento ***A*** como sendo:

$$P(A) = \frac{TAMANHO\ DO\ EVENTO\quad A}{TAMANHO\ DO\ ESPAÇO\ AMOSTRAL\quad S}$$

Exemplo: Lançamento de uma moeda.

Mas como podemos calcular as probabilidades "a priori" nas seguintes situações:

- Uma pessoa que fuma um pacote de cigarros por dia desenvolver cancer.
- Ocorrer uma geada no próximo inverno.
- Haver uma "explosão" na população de desfolhadores presente numa floresta de eucalipto.
- Encontrar um árvore de mogno num parcela de 1ha numa floresta nativa.
- A produção média de uma floresta superar $400m^3$/ha.

CONCEITO FREQUENTISTA OU "A POSTERIORI"

A probabilidade dos eventos é definida com base em **registros da ocorrência** dos eventos.

Os "experimentos aleatórios" passam a ser ***ensaios conduzidos repetidamente em condições relativamente uniformes***, daí o nome de probabilidade "a posteriori".

A probabilidade de um evento é definida como a ***freqüência relativa*** de ocorrência do evento num série de ensaios ou simulações, daí o nome "frequentista".

- A única forma de se conhecer as intensões de voto da população brasileira é entrevistando uma amostra dela. É impossível saber a opinião das pessoas "a priori".
- Numa floresta nativa ou plantada, as únicas informações que dispomos são provenientes de levantamentos (amostras).

Axiomas de Probabilidade

Independentemente do conceito utilizado, clássico ou freqüêntista, o modelo de probabilidade em Bioestatística terá sempre uma **COERÊNCIA INTERNA** que resulta dos AXIOMAS de probabilidade:

Axioma I

$$0 \leq P(A) \leq 1$$

Axioma II

$$P(S) = 1$$

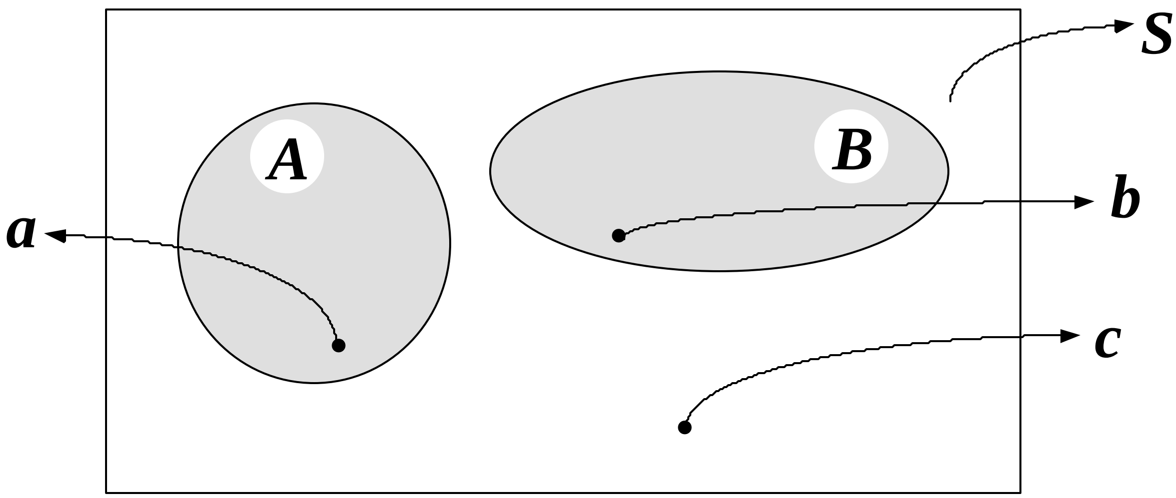
$$P(\varnothing) = 0$$

CONCEITO CLÁSSICO ⇒ P(A) = Tamanho A / Tamanho S
⇒ P(A) tem que ser necessariametne um número entre 0 e 1.

CONCEITO FREQÜENTISTA	⇒ P(A) é a freqüência relativa de A ⇒ P(A) é necessariamente um número entre 0 e 1.

Diagramas de Venn

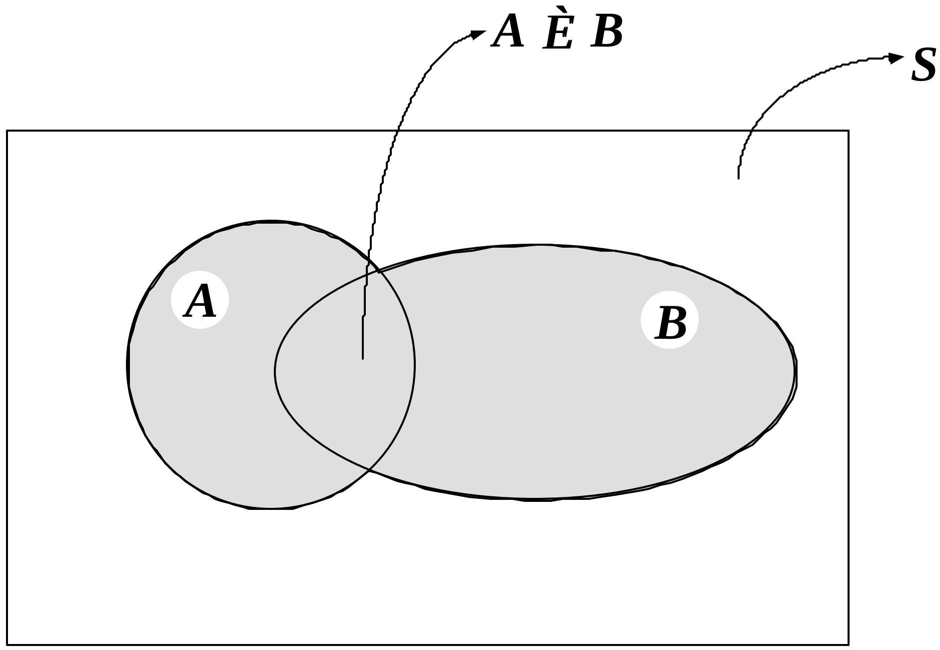
Os diagramas de Venn são utilizados para se "visualizar" os conceitos de básicos de probabilidade.



a, b, c = resultados   A, B = eventos   S = espaço amostral

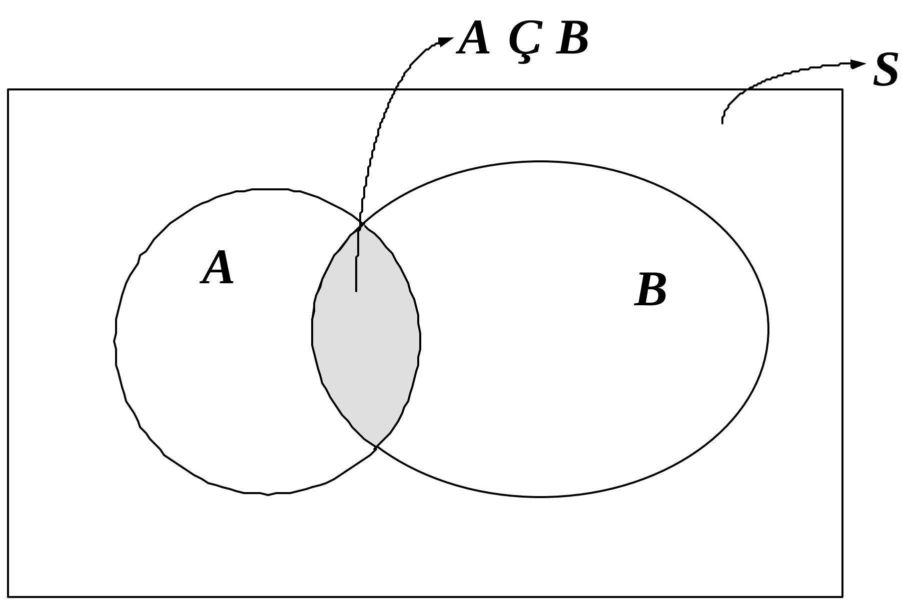
UNIÃO DE 2 EVENTOS

Dizemos:	"ocorre A **ou** B"
Notação:	$A \cup B$



INTERSECÇÃO DE 2 EVENTOS

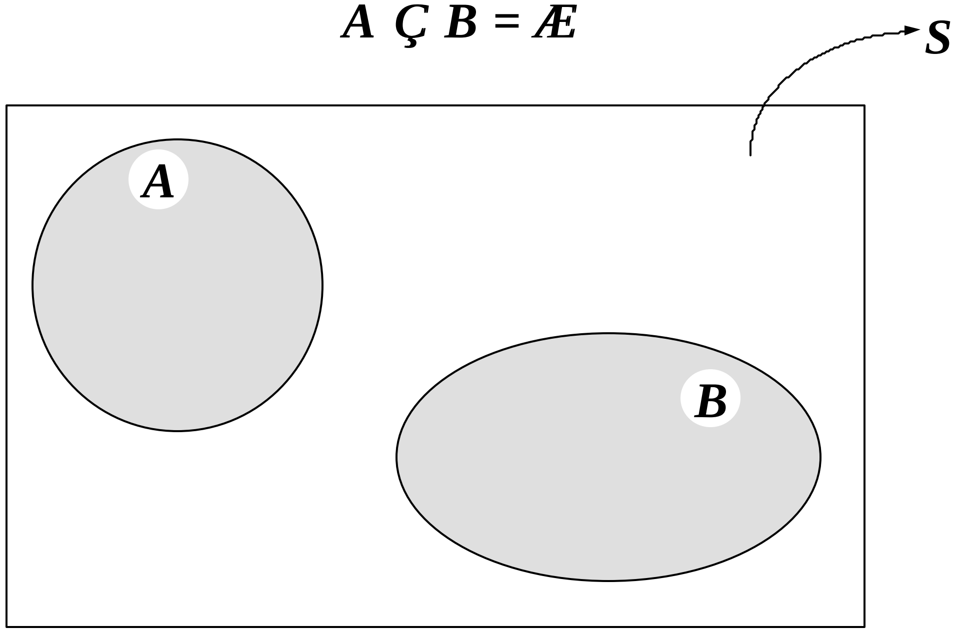
Dizemos: "ocorre A **e** B"
Notação: $A \cap B$



EVENTOS DISJUNTOS:

Dizemos: "A e B são **disjuntos**" ou "A e B são mutuamente exclusivos"
Notação: $A \cap B = \varnothing$

A Ç B = Æ

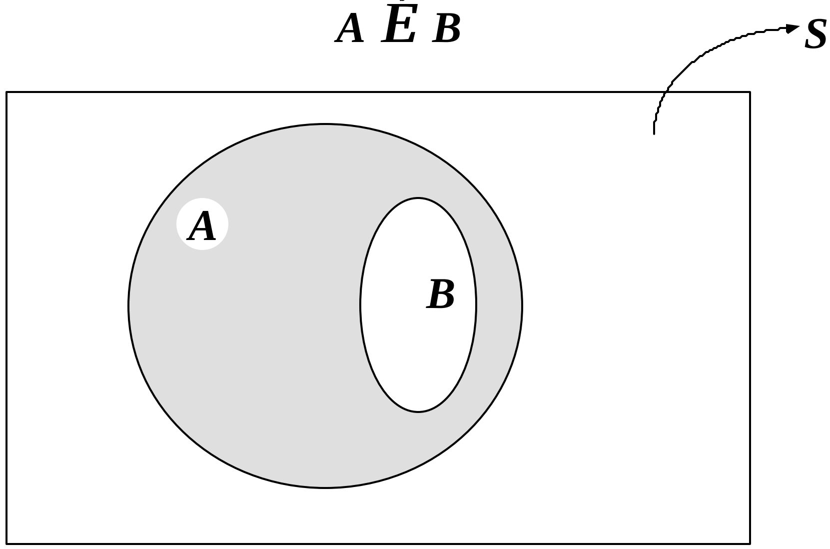


SUB-CONJUNTOS:

Dizemos: "B é **sub-conjunto** de A" ou "B é **implica** em A" .
Notação:

$$B \subset A \quad \Rightarrow \quad B \cap A = B$$
$$\Rightarrow \quad B \cup A = A$$

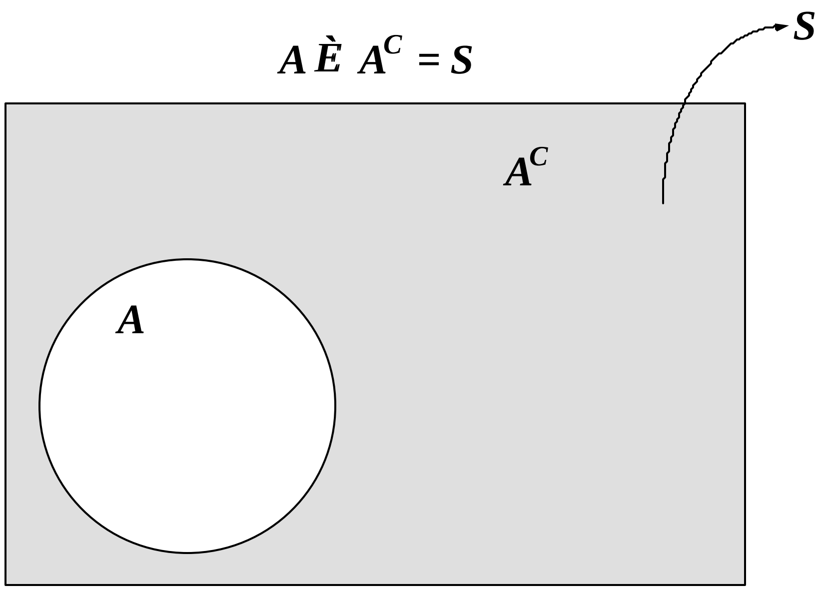
A É B



COMPLEMENTO:

Dizemos: "A^C (o complemento de A) é a **negação** de A".
Notação: $A^C \Rightarrow A^C \cap A = \varnothing$

$A^C \Rightarrow A^C \cup A = S$


A ∪ A^C = S
S
A^C
A

Exemplo 7.E: Lançamento de Dois Dados

Vejamos como podemos utilizar os conceitos apresentados nos diagramas de Venn no caso de um experimento aleatório que consiste no lançamento de dois dados justos:

A = “pelo menos um 3”
B = “soma dos número é no máximo 6”
C = “soma é no mínimo 10”
D = “exatamente dois 3”
E = “nenhum 3”

1. $P(A \bigcup B) =$____________________ 2. $P(A \bigcap B) =$_____________________

3. $P(B \bigcap C) =$____________________ 4. $P(B \bigcup C) =$_____________________

5. $P(A \bigcap D) =$____________________ 6. $P(A \bigcup D) =$_____________________

7. $P(A \bigcap E) =$____________________ 2. $P(A \bigcup E) =$_____________________

Regras de Cálculo de Probabilidades

Utilizando os diagramas de Venn se torna mais fácil compreender algumas regras que surgem naturalmente no cálculo de probabilidades.

REGRA 1:
PROBABILIDADE DA UNIÃO DE EVENTOS

$$P(A \bigcup B) = P(A) + P(B) - P(A \bigcap B)$$

Exemplo 7.F: Lançamento de Dois Dados - II

No lançamento de dois dados justos temos: A = “a soma é impar” e B = “pelo menos um número 1”.
Qual a probabilidade da soma ser impar ou termos pelo menos um número 1?

$$\left.\begin{array}{r} P(A) = 18/36 \\ P(B) = 11/36 \\ P(A \bigcap B) = 6/36 \end{array}\right\} \Rightarrow P(A \bigcup B) = \frac{18+11-6}{36} = \frac{23}{36}$$

REGRA 1B:
PROBABILIDADE DA UNIÃO DE EVENTOS DISJUNTOS

$$A \text{ e } B \text{ disjuntos} \Rightarrow A \cap B = \varnothing \quad \Rightarrow P(A \cap B) = 0$$
$$\Rightarrow P(A \cup B) = P(A) + P(B)$$

Exemplo 7.G: Lançamento de Dois Dados - III

No lançamento de dois dados justos temos: A = "a soma é menor do que 6" e B = "a soma é no mínimo 10".

$$\left.\begin{array}{r} P(A) = 15/36 \\ P(B) = 6/36 \\ P(A \cap B) = 0 \end{array}\right\} \Rightarrow P(A \cup B) = \frac{15}{36} + \frac{6}{36} = \frac{21}{36}$$

Desigualdade de Boole

REGRA 2:
PROBABILIDADE DA DE UMA SEQÜÊNCIA DE EVENTOS DISJUNTOS

Se $A_1, A_2, A_3, \ldots,$ formam uma seqüência de eventos disjuntos, então

$$P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

REGRA 3:
DESIGUALDADE DE BOOLE

Dada uma seqüência de eventos $A_1, A_2, A_3, \ldots, A_n,$

$$P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$$

Exemplo 7.H: Deficiência Nutricional

Numa floresta de *Eucalyptus grandis*, foram coletadas amostras de folhas de 100 árvores (selecionadas aleatoriamente) para análise foliar. A análise revelou deficiência de alguns nutrientes.

Nutrientes		Proporção de Árvores c/ Deficiência
Macro	N	0.35
	P	0.15
	K	0.10
Micro	B	0.50
	Zn	0.01
	Mn	0.02

Qual a probabilidade de uma árvore selecionada aleatoriamente nesta floresta estar com deficiência de macro-nutrientes?

0.35 + 0.15 + 0.10 = 0.60

A Desigualdade de Boole implica que a probabilidade é de no **máximo 60%,** isto é, que no **mínimo 40%** das árvores não tem deficiência de macro-nutrientes.

Probabilidade do Complemento

No lançamento de duas moedas temos: A = "pelo menos uma cara", B = "duas coroas".
Qual a probabilidade duas coroas ou pelo menos uma cara ?

Moeda 1	Moeda 2: C	Moeda 2: K
C	CC	CK
K	KC	KK

$$\left.\begin{array}{r} P(A) = 3/4 \\ P(B) = 1/4 \\ P(A \cap B) = 0 \end{array}\right\} \Rightarrow P(A \cup B) = \frac{3}{4} + \frac{1}{4} = \frac{4}{4} = 1$$

$$A \text{ e } B \text{ são complementares } \Rightarrow P(A) + P(B) = P(S) = 1$$

REGRA 4:
PROBABILIDADE DO COMPLEMENTO

$$P(S) = P(A) + P(A^C) = 1 \Rightarrow P(A) = 1 - P(A^C)$$

Exemplo 7.I: Lançamento de Um Dado

Um dado é lançado 10 vêzes, qual a probabilidade de A = "pelo menos um 6"?

A^C = "nenhum 6" $= \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} * \frac{5}{6} = \frac{5^{10}}{6^{10}} = \frac{9765625}{60466175} = 0.1615$

$P(A) = 1 - P(A^C) = 1 - 0.1615 = 0.8385$

Probabilidade Condicional

PROBABILIDADE CONDICIONAL

A ocorrência de um evento **altera** a probabilidade de ocorrência de outro evento

Não estabelece relação de **causa-efeito** entre os eventos, simplesmente relaciona as probabilidades dos eventos ocorrerem.

Exemplo 7.J: Educação em Sala

Num estudo de educação, alunos do primeiro e segundo graus foram questionados sobre atividades de sua preferência dentro de sala de aula.

Um aluno é selecionado aleatoriamente deste grupo:

Qual a probabilidade dele ser do 1o. grau?

$$P(A) = \frac{30}{50} = \frac{3}{5}$$

Qual a probabilidade dele preferir atividades artísticas?

$$P(B) = \frac{32}{50} = \frac{16}{25}$$

Se o aluno selecionado for do 1o. grau, qual a probabilidade dele preferir atividades artísticas?

$$P(B \text{ dado que ocorreu } A) = \frac{27}{30} = \frac{9}{10}$$

Se o aluno preferir atividades artísticas, qual a probabilidade dele ser do 1o. grau?

$$P(A \text{ dado que ocorreu } B) = \frac{27}{32}$$

GRAU	ATIVIDADE		
	ARTÍSTICA	LEITURA	TOTAL
1O.	27	3	30
2O.	5	15	20
TOTAL	32	18	50

NOTAÇÃO

$$P(A \text{ dado que ocorreu } B) = P(A|B)$$

$$P(B \text{ dado que ocorreu } A) = P(B|A)$$

REGRA 5:
PROBABILIDADE CONDICIONAL

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Exemplo 7.J: Educação em Sala

Voltando ao exemplo de educação temos:

- A = "aluno do primeiro grau" $\Rightarrow P(A) = 3/5$
- B = "aluno prefere atividades artísticas" $\Rightarrow P(B) = 16/25$
- A e B = "aluno do primeiro grau e prefere atividades artísticas" $\Rightarrow P(A \cap B) = 27/50$

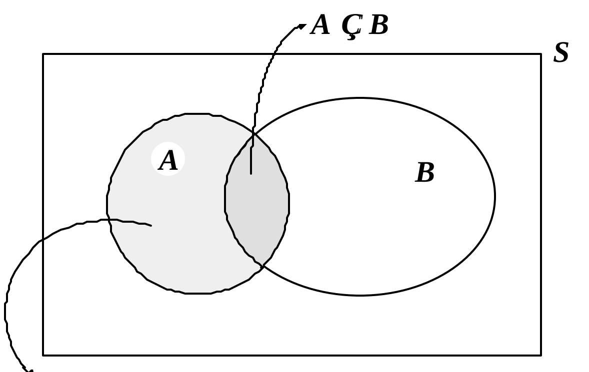
A dado B = "aluno do primeiro grau dado que o aluno prefere atividades artísticas"

$$\Rightarrow P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{27/50}{16/25} = \frac{27}{32}$$

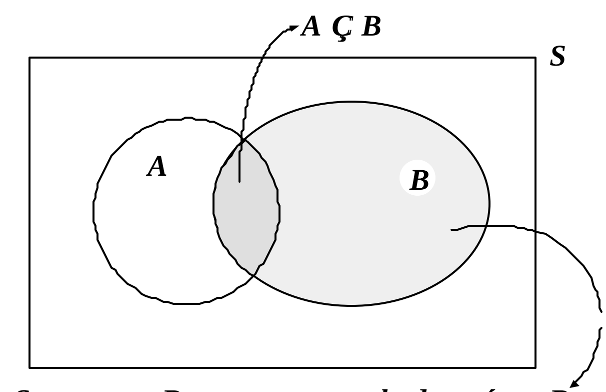
B dado A = "aluno prefere atividades artísticas dado que o aluno é do primeiro grau"

$$\Rightarrow P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{27/50}{3/5} = \frac{27}{30} = \frac{9}{10}$$

Probabilidade Condicional no Diagrama de Venn



Se o evento A ocorreu, o resultadoestá em A.
P(B|A) = P(A Ç B)/P(A)



Se o evento B ocorreu, o resultado está em B.
P(A|B) = P(A Ç B)/P(B)

Exemplo 7.K: Concorrência

Uma empresa de consultoria participa de duas concorrências para realizar estudos de impacto ambiental. A probabilidade dela vencer a primeira concorrência é de 50% e de vencer a segunda é de 70%, enquanto que a probabilidade de vencer ambas concorrências é 40%.

- A = vence a 1a. concorrência $\Rightarrow P(A) = 0.5$
- B = vence a 2a. concorrência $\Rightarrow P(B) = 0.7$
- A e B = vence ambas as concorrências $\Rightarrow P(A \cap B) = 0.4$

Qual a probabilidade de vencer a segunda concorrência dado que ela vence a primeira?

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.4}{0.5} = 0.80$$

Qual a probabilidade de vencer a primeira concorrência dado que ela vence a segunda?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.4}{0.7} = 0.57$$

Probabilidade da Intersecção de Dois Eventos

A probabilidade condicional nos permite calcular diretamente a probabilidade da intersecção de dois eventos.

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(A \cap B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A \cap B) = P(A|B)P(B)$$

REGRA 6:
PROBABILIDADE DA INTERSECÇÃO DE DOIS EVENTOS

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

Exemplo 7.L: Seleção de Pessoal

Uma empresa produtora de papel e celulose dispõe de 250 fichas cadastrais de candidatos a vagas de emprego. Assume-se que as fichas cadastrais representem uma amostra aleatória da população da economicamente ativa cidade. Nas fichas, 60% são homens e 40% são mulheres. Sabe-se que nesta cidade 50% dos homens são fumantes, mas apenas 20% das mulheres são fumantes.

Qual a proporção da população que qualificaria para um emprego de motoserrista (homem não fumante)?

- A = homem: $P(A) = 0.6$
- B = homen não fumante: $P(B|A) = 0.5$

$$P(A \cap B) = P(B|A)P(A) = (0.5)(0.6) = 0.30$$

Qual a proporção da população que qualificaria para um emprego de embaladora (mulher não fumante)?

- A = mulher: $P(A) = 0.4$
- B = mulher não fumante: $P(B|A) = 0.8$

$$P(A \cap B) = P(B|A)P(A) = (0.4)(0.8) = 0.32$$

Independência de Eventos

Dois eventos são ditos **INDEPENDENTES** quando:	⇨	*A ocorrência de um dos eventos não altera a probabilidade de ocorrência do outro.*

DEFINIÇÃO:
DOIS EVENTOS INDEPENDENTES

$$P(A|B) = P(A)$$
$$P(B|A) = P(B)$$

Exemplo 7.M: Amostragem Sem Reposição x Com Reposição

Uma urna contém 5 bolas: 2 brancas e 3 pretas. Duas bolas são selecionadas aleatoriamente.

A = a primeira bola é preta $\Rightarrow P(A) = 3/5$

B = a segunda bola é branca. $\Rightarrow P(B) = 2/5$

- Se bolas forem retiradas **sem reposição** qual á probabilidade de B dado A ?

Se A ocorreu, uma bola preta foi tirada da urna e, portanto, restaram 2 brancas e 2 pretas, logo:

$$P(B|A) = \frac{2}{4} = \frac{1}{2} \neq P(B) = \frac{2}{5}$$

Os eventos A e B não são independentes.

- Se bolas forem retiradas **com reposição** qual á probabilidade de B dado A ?

Se A ocorreu, mas a bola preta retornou a urna, a configuração da urna permanece 2 brancas e 3 pretas, logo

$$P(B|A) = \frac{2}{5} = P(B) = \frac{2}{5}$$

Os eventos A e B são independentes.

REGRA 7:
PROBABILIDADE DA INTERSECÇÃO DE DOIS EVENTOS INDEPENDENTES

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

Exemplo 7.N: Cruzamento Aleatório

Uma certa planta anual, possui um gene com dois alelos: o dominante (***A***) e o recessivo (***a***). Como cada indivíduo possui dois alelos, os genótipos possíveis numa população são ***AA***, ***Aa*** e ***aa***. Numa dada população, as proporções destes genótipos são: 1/2, 1/4 e 1/4, respectivamente. Assumindo que os indivíduos desta população se cruzem de modo totalmente aleatório, quais as proporções dos genótipos na próxima geração?

Genótipo	Criação dos Gametas pelos Genótipos	Proporçao de Gametas	Genótipos na Próxima Geração
AA	A => (1/2) * (1/2) = 1/4 A => (1/2) * (1/2) = 1/4	A => 5/8	AA => (5/8)*(5/8) = 25/8
Aa	A => (1/4) * (1/2) = 1/8 a => (1/4) * (1/2) = 1/8	a => 3/8	Aa => 2*(5/8)*(3/8) = 30/8
aa	a => (1/4) * (1/2) = 1/8 a => (1/4) * (1/2) = 1/8		aa => (3/8)*(3/8) = 9/8

Exemplo 7.O: Ocorrência de Paranormalidade

Os fenômenos paranormais são invensões de pessoas supersticiosas ou realmente existem?

SITUAÇÃO: Uma pessoa sonha que um amigo morreu e nessa mesma noite o amigo morre.
PRESSUPOSIÇÕES: a) sonho e morte são independentes (não há paranormalidade)
b) cada pessoa tem apenas um sonho destes na vida.

Evento A_i
Pessoa i tem o sonho num dado dia.
Expectativa de vida do brasileiro:
65 anos = 23750 dias

$$P(A_i) = \frac{1}{23750}$$

Evento B_i
O amigo da pessoa i morre num dado dia. Taxa de mortalidade de 1%/ano: 1.700.000 (0.01) = 1700000 pessoas/ano
= 4658 pessoas/dia

$$P(B_i) = \frac{4658}{170000000}$$

<u>Ocorrência de ambos eventos num dado dia assumindo independência</u>:

$$P(A_i \cap B_i) = P(A_i)P(B_i) = \left(\frac{1}{23750}\right)\left(\frac{4658}{170000000}\right) = 1.1537 \times 10^{-9}$$

Conclusão
o fenômeno $(A_i \cap B_i)$
é extremamente raro!!

<u>Mas a população é grande, será que a ocorrência simultânea ainda é rara na população?</u>

Conclusão
1 vez a cada 5 dias o fenômeno $(A_i \cap B_i)$ ocorre!!

Conceitos-Chave

SIMULAÇÃO - EXPERIMENTO ALEATÓRIO - ESPAÇO AMOSTRAL - EVENTO - CONCEITO CLÁSSICO DE PROBABILIDADE - PROBABILIDADE "A PRIORI" - CONCEITO FREQÜENTISTA DE PROBABILIDADE - PROBABILIDADE "A POSTERIORI" - AXIOMAS DE PROBABILIDADE - DIAGRAMS DE VENN - UNIÃO DE EVENTOS - INTERSECÇÃO DE EVENTOS - EVENTOS DISJUNTOS - EVENTOS MUTUAMENTE EXCLUSIVOS - SUB-CONJUNTOS - COMPLEMENTO - PROBABILIDADE DA UNIÃO - PROBABILIDADE DA SEQÜÊNCIA DE EVENTOS DISJUNTOS - DESIGUALDADE DE BOOLE - PROBABILIDADE DO COMPLEMENTO - PROBABILIDADE CONDICIONAL - PROBABILIDADE DA INTERSECÇÃO - EVENTOS INDEPENDENTES - PROBABILIDADE DA INTERSECÇÃO DE EVENTOS INDEPENDENTES

Exercícios

7.1 Numa floresta nativa a taxa de ocorrência de espécies arbóreas pioneiras é de 20%. Através de simulações (utilizando a tabela de números aleatórios), estime a probabilidade de numa amostra aleatória de 10 árvores se encontrar no mínimo três árvores de espécies pioneiras.

7.2 Esquematize um processo de simulação (utilizando números aleatórios) para a amostragem de árvores numa floresta nativa onde:

- 10% das árvores são de espécies pioneiras;
- 20% das árvores são de espécies secundárias iniciais;
- 30% das árvores são de espécies secundárias tardias;
- 40% das árvores são de espécies primárias.

O objetivo é simular uma amostra aleatória de 20 árvores desta floresta.

7.3 Uma urna contém três bolas: uma azul (A), uma verde (V) e uma rosa (R). Duas bolas são selecionadas aleatoriamente da urna. O resultado é representado pela combinação da letra das cores, por exemplo, se as bolas selecionadas forem a azul e a verde o resultado é (A,V). Represente o espaço amostral para este experimento nas seguintes situações:

a) Amostragem <u>com reposição</u> onde a ordem é importante;
b) Amostragem <u>com reposição</u> onde a ordem <u>não</u> é importante;
c) Amostragem <u>sem reposição</u> onde a ordem é importante;
d) Amostragem <u>sem reposição</u> onde a ordem <u>não</u> é importante.

7.4 Num levantamento da avifauna, as aves foram classificadas segundo a sua dieta preferencial em insentívoras (I) e frugívoras (F).

a) Represente o espaço amostral para uma amostra de três aves;
b) Represente o evento A = "pelo menos uma ave é frugívora";
c) Represente o evento B = "exatamente duas aves são frugívoras".

7.5 Considere com experimento aleatório o lançamento de dois dados balanceados.

a) Represente o espaço amostral;
b) Represente o evento A = "exatamente 1 dado é três";
c) Represente o evento B = "exatamente 2 dados são três";
d) Represente o evento C = "pelo menos um dos dados é três";
e) Represente o evento D = "a soma dos dados é seis".

7.6 Considere com experimento aleatório o lançamento de dois dados justo. Calcule a probabilidade dos seguintes eventos:

A. = “nenhum três”;
B. = “exatamente um três”;
C. = “exatamente dois três”;
D. = “pelo menos um três”;
E. Compare 1 - P(A) com P(D).

7.7 Num levantamento em floresta de *Pinus oocarpa*, foram observadas 830 árvores, segundo a tabela abaixo. Os defeitos são excludentes, isto é, cada árvore foi classificada em apenas uma das quatro classes de defeito.

	Classes de Defeito do Tronco				
Idade	**Bifurcada**	**Tortas**	**Rabo-de-raposa**	**Sem Defeito**	**Total**
Jovem	24	91	78	181	374
Madura	36	74	76	270	456
Total	60	165	154	451	830

Considere os seguintes eventos:
A. = “árvore jovem”;
B. = “árvore madura”;
C. = “árvore bifurcada”;
D. = “árvore torta”;
E. = “árvore com rabo-de-raposa”;
F. = “árvore jovem e bifurcada”;
G. = “árvore madura e torta”;
H. = “árvore jovem sem defeitos”.

Estabeleça a relação entre os seguintes eventos:
a) A e B;
b) A e G;
c) C e F;
d) D e E;
e) $B \cap D$ e G**;**
f) $A \cap (C \cup D \cup E)$ e H.

7.8 Considerando as informações do exercício anterior calcule as seguintes probabilidades:
a) $P(A \cup C)$
b) $P(B \cup E)$
c) $P(D \cup F)$
d) $P(C \cup G)$
e) $P(B \cup H)$
f) $P(C \cup D \cup E)$
g) $P(A \cap (C \cup D \cup E))$

7.9 Uma indústrias de móveis de *Pinus* fez um levantamento no pátio de secagem para analisar a incidência de defeitos nas peças de madeira secas ao ar. Todos os defeitos de cada peça amostrada (amostra aleatória em 3 lotes) foram anotados, isto é, uma mesma peça pode ter mais de um defeito. O resultado encontrado foi.

Defeitos	**Freqüência Relativa**		
	Lote 1	**Lote 2**	**Lote 3**

Manchamento	0.23	0.19	0.35
Empenamento Longitudinal	0.07	0.03	0.12
Empenamento Transversal	0.09	0.03	0.10
Nós Mortos	0.01	0.02	0.05
Perfuração por Brocas	0.02	0.05	0.20
Rachaduras de Topo	0.17	0.20	0.25
Madeira Juvenil	0.08	0.08	0.17

a) Qual a proporção de peças em cada lote que apresentam defeitos de empenhamento ? Por que?
b) Qual dos lotes está em melhores condições, isto é, apresenta a menor proporção defeituosa ? Por que?

7.10 Um estudo do coportamento social de capivaras produziu o seguinte resultado:

	Comportamento		
Ambiente	Agressivo	Não Agressivo	Total
Restrito	22	5	27
Amplo	20	12	32
Total	42	17	59

Se um animal for selecionada aleatoriamente deste grupo:
a) Qual a probabilidade deste animal estar num ambiente restrito?
b) Qual a probabilidade deste animal ter um comportamento agressivo?
c) Se este animal está num ambiente restrito, qual a probabilidade dele ter um comportamento agressivo?
d) Se este animal está num ambiente restrito, qual a probabilidade dele ter um comportamento não agressivo?
e) Se este animal tem um comportamento agressivo, qual a probabilidade dele estar num ambiente restrito?
f) Se este animal tem um comportamento agressivo, qual a probabilidade dele estar num ambiente amplo?

7.11 Um aluno de Engenharia Florestal considera as chances de conseguir dois estágios prático. As chances dele conseguir o estágio numa empresa florestal são de 80%, enquanto que as chances de conseguir um estágio num parque nacional são de 70%. As possibilidades de conseguir ambos os estágios, entretanto, são de 50%. Pergunta-se:
a) Qual a probabilidade do aluno conseguir o estágio no parque nacional, dado que ele conseguiu o estágio na empresa?
b) Qual a probabilidade do aluno conseguir o estágio na empresa florestal, dado que ele conseguiu o estágio no parque nacional?
c) Qual a probabilidade do aluno conseguir pelo menos um dos estágios?

7.12 Uma fábrica de peças possui duas linhas de produção. A linha A produz 5% das peças da fábrica e destas 10% são defeituosas. Já a linha B produz 95% das peças com uma taxa de defeitos de apenas 2%.
a) Qual a proporção de peças defeituosas que deixam a fábrica?
b) Se a peça é defeituosa, qual a probabilidade dela ter saido da linha de produção A?
c) Se a peça é defeituosa, qual a probabilidade dela ter saido da linha de produção B?
d) Se a peça **não** é defeituosa, qual a probabilidade dela ter saido da linha de produção A?
e) Se a peça **não** é defeituosa, qual a probabilidade dela ter saido da linha de produção B?

7.13 Numa floresta de Pinus elliottii, 30% das árvores foram resinadas. Das árvores não resinadas, 70% são apropriadas para serraria, enquanto que dentre as resinadas apenas 10% o são. Assumindo que uma árvore desta floresta for selecionada ao acaso, pegunta-se:

a) Qual a probabilidade dela ser apropriada para serraria?
b) Qual a probabilidade dela ter sido resinada e ser própria para serraria?
c) Qual a probabilidade dela ter sido resinada e não ser própria para serraria?
d) Qual a probabilidade dela não ter sido resinada e ser própria para serraria?

7.14 Sabe-se que numa região de florestas tropicais em condições primitivas 75% das árvores são de espécies raras, mas em áreas que sofreram distúrbios antrópicos a proporção de espécies raras cai para 30%. Considere uma reserva florestal onde 20% da sua área foi alterada antropicamente. Qual a probabilidade de um árvores selecionada aleatoriamente nesta reserva ser de espécie rara ?

7.15 Um dado é lançado 20 vêzes, qual a probabilidade de pelo menos um dos lançamentos resultar num seis?

7.16 Numa floresta de *Eucalyptus grandis* a taxa de ocorrência de cancro é de 2.5%. Num levantamento foram selecionadas ao acaso 30 árvores, qual a probabilidade de pelo menos uma ter cancro?

7.17 Uma festa reune 23 pessoas. Assumindo que as datas de nascimento das pessoas ocorre de modo totalmente aleatório, qual a probabilidade de pelo menos duas pessoas terem nascido no mesmo dia do ano?

7.18 Sabe-se que as árvores de uma certa espécie arbórea produzem em média 2000 frutos no estágio inicial de desenvolvimento. Quando os frutos amadurecem, aproximadamente 27 sementes por fruto são disseminadas. Outras informações sobre a biologia reprodutiva desta espécie incluem:

- 55% dos frutos em início de desenvolvimento são abortados.
- Dos frutos verdes (não abortados), 70% são predados antes de amadurecer.
- Dos frutos que amadurecem, 50% das sementes são eficientemente disseminadas.

Qual a quantidade de sementes por árvore que auxiliará efetivamente na regeneração natural desta espécie.

7.19 Num sistema de iluminação, foram instaladas em série 15 lâmpadas, de modo que se uma delas queimar todas as lâmpadas apagam. Cada lâmpada tem probabilidade de 0.99 de não queimar nas primeiras 300 horas de funcionamento. Assumindo que as lâmpadas se comportam de modo independente, qual a probabilidade do sistema não falhar nas primeiras 300 horas ?

7.20 Um lote de sementes de *Eucalyptus saligna* com uma proporção de 5% de sementes híbridas (*E. saligna* **x** *E. cloeziana*) foi utilizado para a implantação de uma floresta. Se dez árvores desta floresta forem selecionadas ao acaso qual a probabilidade de:

a) nenhuma delas ser híbrido;
b) pelo menos uma delas ser híbrido;
c) todas elas serem híbridos.

11. O Modelo Fundamental: "Distribuição Normal"

Introdução

As principais razões que fazem a distribuição Normal o modelo mais importante na Bioestatística são:

1. Muitas variáveis biométricas tendem a ter distribuição Normal. Isto ocorre principalmente quando a variável é influenciada por um grande número de fatores que atuam de modo independente e aditivo.
2. A distribuição das médias amostrais de uma variável qualquer tendem a ter distribuição Normal, mesmo que a variável em si não tenha distribuição Normal.
3. Muitos testes e modelos estatísticos têm como pressuposição a "nomalidade dos dados", isto é, que os dados possuem distribuição Normal.

Histórico

A distribuição Normal é tambem conhecida como "distribuição Gaussiana" como homenagem a Karl F. Gauss (1777-1855), brilhante matemático e físico alemão, que a desenvolveu no início do século XIX. Entretanto, Abraham de Moivre (1667-1754) foi o primeiro a anunciar a equação da distribuição em 1733 e Pierre-Simon Marquis de Laplace (1749-1827), famoso matemático e físico francês, a redescobriu na mesma época que Gauss. Para evitar "uma questão internacional de originalidade" o famoso estatístico inglês Karl Pearson passou a chamá-la de distribuição "Normal" em 1920.

Definição da Distribuição Normal

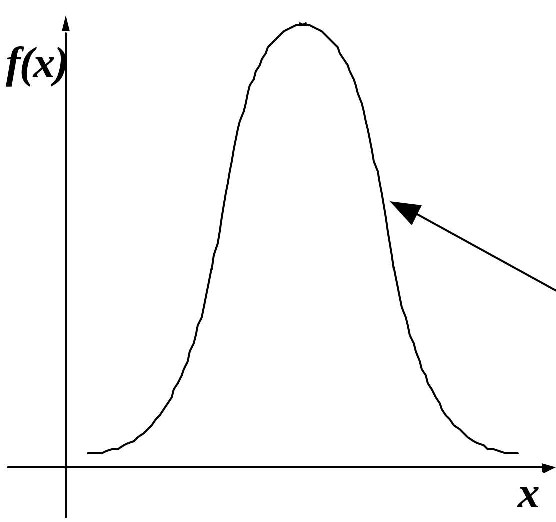
A Distribuição Normal
é uma
Variável Aleatória Contínua

É definida por uma
função de densidade

Função de Densidade da Distribuição Normal

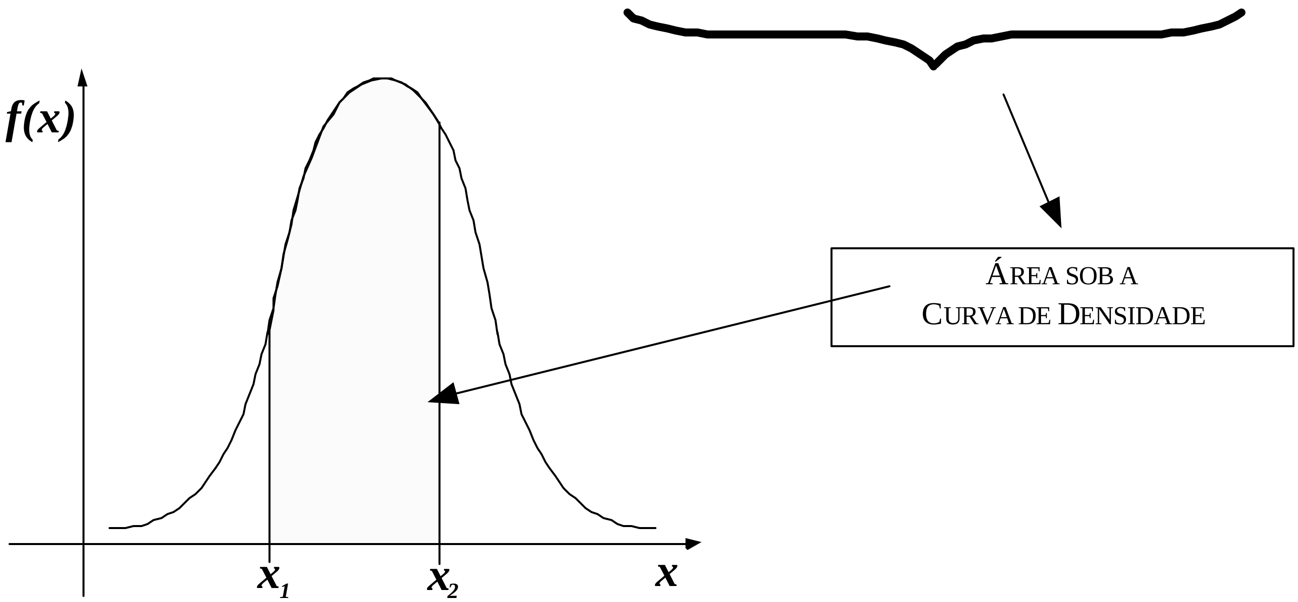
$$f(x) = \left(\frac{1}{\sqrt{2p}} \right) \frac{1}{s} e^{-\frac{1}{2}\left[\frac{x-m}{s} \right]^2}$$

Define a
forma da distribuição



Probabilidade de $\boldsymbol{x}$ estar entre $\boldsymbol{x_1}$ e $\boldsymbol{x_2}$:

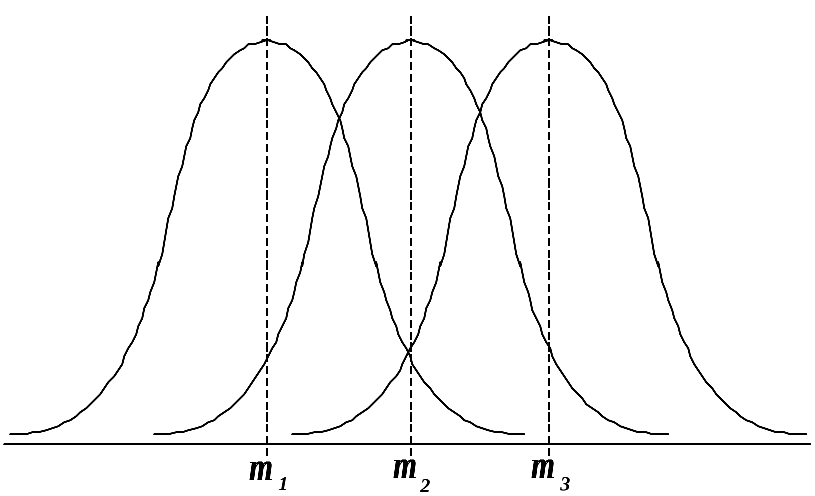
$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(t)dt = \int_{x_1}^{x_2} \left(\frac{1}{\sqrt{2p}}\right)\frac{1}{s}\,\mathrm{e}^{-\frac{1}{2}\left[\frac{t-m}{s}\right]^2} dt$$



Propriedades da Distribuição Normal

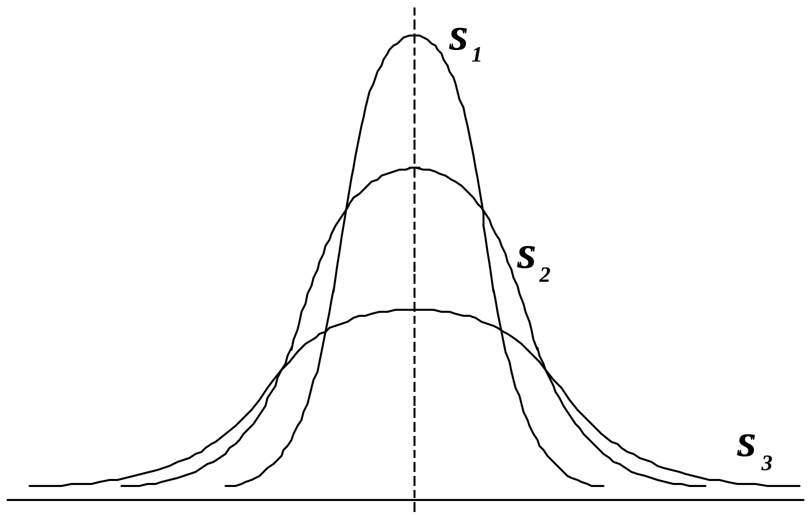
1) <u>FORMA DE "SINO"</u>: unimodal e simétrica

2) <u>DOIS PARÂMETROS</u>: média ($\boldsymbol{m}$) e desvio padrão ($\boldsymbol{s}$)



A média ($\boldsymbol{m}$) controla a localização do centro da distribuição, é o ponto de simetria.

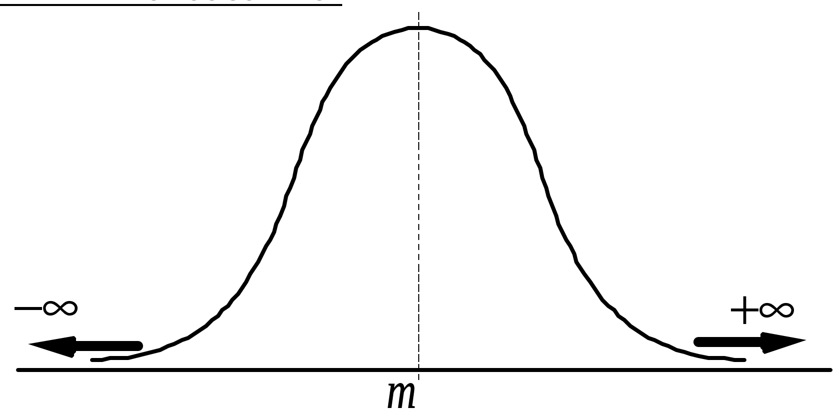
$$m_1 < m_2 < m_3$$



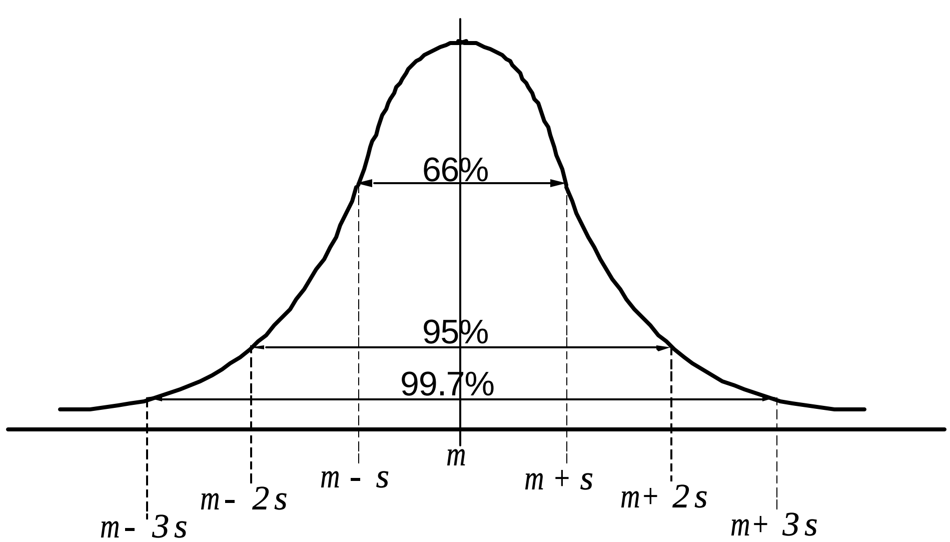
O desvio padrão (s) controla a dispersão da curva ao redor da média.

$s_1 < s_2 < s_3$

3) <u>NÃO POSSUI LIMITE INFERIOR OU SUPERIOR</u>



4) <u>UNIDADES PADRÕES</u>: o desvio padrão define “unidades padrões” na distribuição a partir da média, isto é, a dispersão dos dados é controlada pelas “unidades de desvio padrão”.



Distribuição Normal Padronizada

A média (m) e o desvio padrão (s) definem a distribuição Normal.

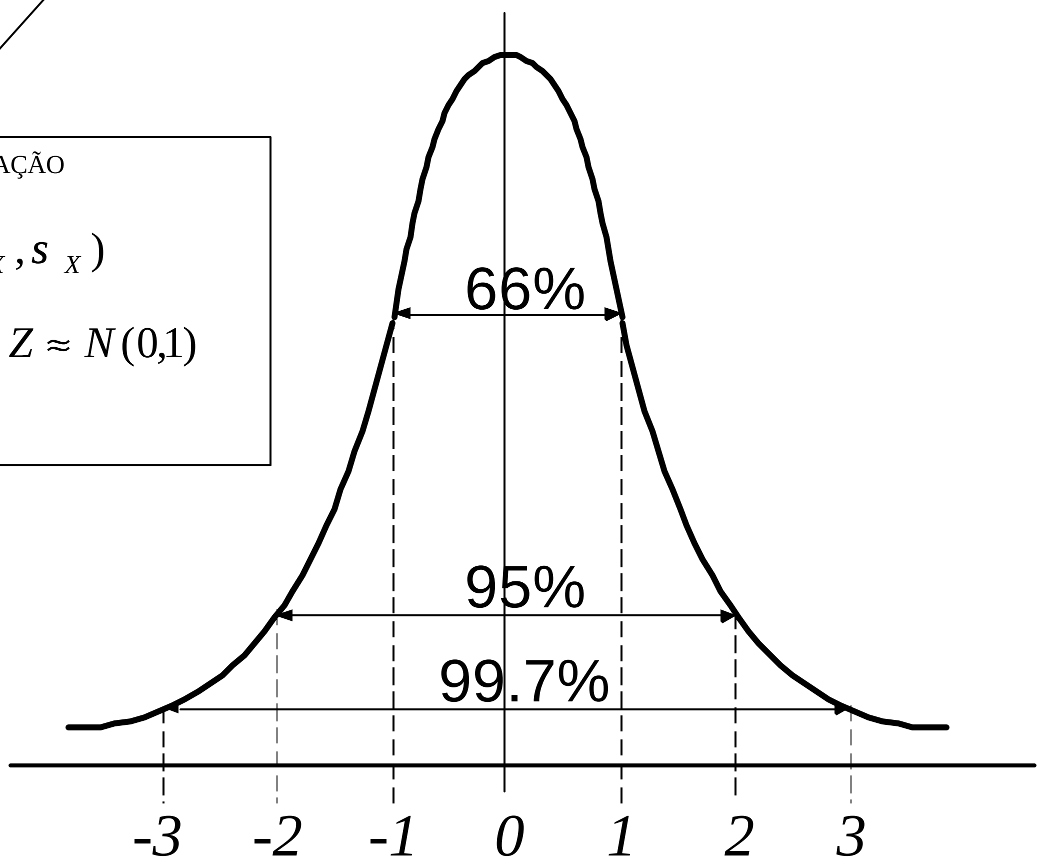
Existem tantas distribuições Normais quantos valores de m e s podemos imaginar (infinitas).

Todas elas podem ser reduzidas a uma única distribuição Normal através da Padronização.

PADRONIZAÇÃO

$$X \approx N(m_X, s_X)$$

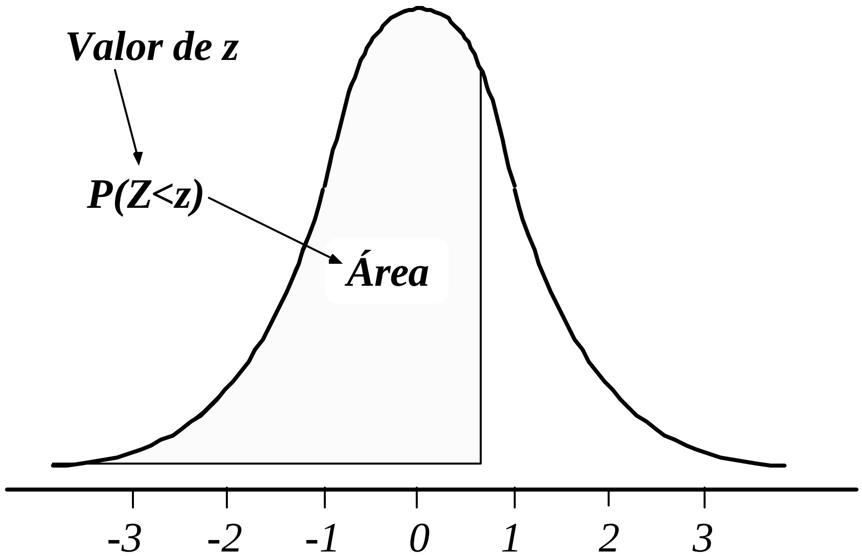
$$Z = \frac{X - m_X}{s_X} \Rightarrow Z \approx N(0,1)$$



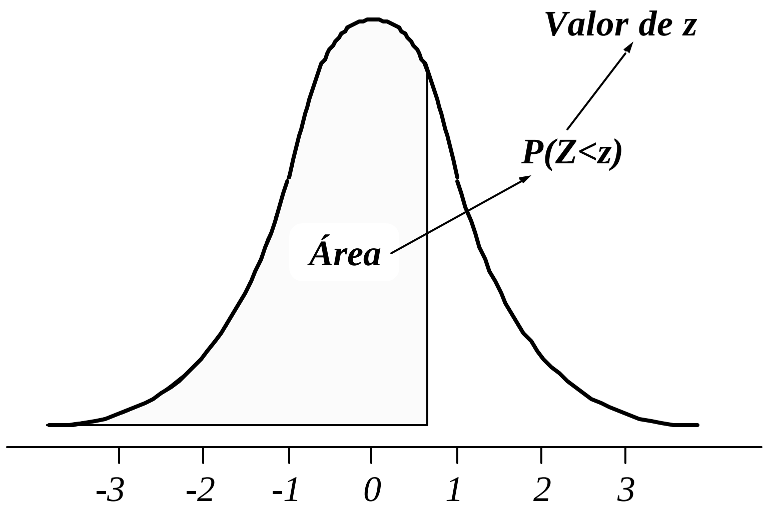
De modo inverso, a Variável Normal Padronizada (Z) pode ser transformada em qualquer Variável Normal:

$$Z \approx N(0,1)$$

$$X = s_X(Z) + m_X \Rightarrow X \approx N(m_X, s_X)$$



z	P(Z<z)	z	P(Z<z)
-3.0	0.0013	0.0	0.5000
-2.9	0.0019	0.1	0.5398
-2.8	0.0026	0.2	0.5793
-2.7	0.0035	0.3	0.6179
-2.6	0.0047	0.4	0.6554
-2.5	0.0062	0.5	0.6915
-2.4	0.0082	0.6	0.7257
-2.3	0.0107	0.7	0.7580
-2.2	0.0139	0.8	0.7881
-2.1	0.0179	0.9	0.8159
-2.0	0.0228	1.0	0.8413
-1.9	0.0287	1.1	0.8643
-1.8	0.0359	1.2	0.8849
-1.7	0.0446	1.3	0.9032
-1.6	0.0548	1.4	0.9192
-1.5	0.0668	1.5	0.9332
-1.4	0.0808	1.6	0.9452
-1.3	0.0968	1.7	0.9554
-1.2	0.1151	1.8	0.9641
-1.1	0.1357	1.9	0.9713
-1.0	0.1587	2.0	0.9772
-0.9	0.1841	2.1	0.9821
-0.8	0.2119	2.2	0.9861
-0.7	0.2420	2.3	0.9893
-0.6	0.2743	2.4	0.9918
-0.5	0.3085	2.5	0.9938
-0.4	0.3446	2.6	0.9953
-0.3	0.3821	2.7	0.9965
-0.2	0.4207	2.8	0.9974
-0.1	0.4602	2.9	0.9981
0.0	0.5000	3.0	0.9987



P(Z<z)	z (percentil)
1.0	-2.33
2.5	-1.96
5.0	-1.64
10.0	-1.28
20.0	-0.84
30.0	-0.52
40.0	-0.25
50.0	0.00
60.0	0.25
70.0	0.52
80.0	0.84
90.0	1.28
95.0	1.64
97.5	1.96
99.0	2.33

Exemplo 11.A: Cálculo de Probabilidades na Distribuição Normal Padronizada

O Anexo B apresenta uma tabela detalhada da Distribuição Normal Padronizada. Utilizando esta tabela encontre as seguintes probabilidades:

a) $P(Z > 1.96) =$

b) $P(Z < 1.96) =$

c) $P(Z < -1.96) =$

d) $P(-2.50 < Z < 2.50) =$

e) $P(0.50 < Z < 1.50) =$

f) $P(Z < z) = 0.75 \Rightarrow z = ?$

g) $P(-z_1 < Z < z_1) = 0.75 \Rightarrow z_1 = ?$

h) $P(0.27 < Z < z_2) = 0.50 \Rightarrow z_2 = ?$

Exemplo 11.B: Aplicando a Distribuição Normal

Assumindo que o Quociente de Inteligência (QI) de crianças de 12 anos pode ser modelado por uma distribuição Normal com média 100 e desvio padrão 16, responda as questões abaixo:

a) Esquematize o gráfico da distribuição do QI.

b) Qual a proporção de crianças com QI acima de 84?

c) Qual a proporção de crianças com QI entre 96 e 120?

d) Qual a proporção de crianças com QI entre 84 e 116?

e) Qual o QI que uma criança deve possuir para estar entre os 1% mais inteligentes da população?

Exemplo 11.C: Aplicando a Distribuição Normal II

A distribuição dos diâmetros de uma floresta de *Pinus caribaea* var. *caribaea* segue distribuição Normal com média 23 cm e desvio padrão 7 cm. Pede-se:

a) Esquematize o gráfico da distribuição.

b) Qual a proporção de árvores com diâmetro acima de 28cm?

c) Qual a proporção de árvores com diâmetro abaixo de 17cm?

d) Qual a proporção de árvores com diâmetro acima de 20cm?

e) Qual a proporção de árvores com diâmetro abaixo de 30cm?

f) Qual a proporção de árvores com diâmetro entre 20 e 25cm?

g) Qual a proporção de árvores com diâmetro entre 16 e 30cm?

h) Qual a proporção de árvores com diâmetro entre 25 e 30cm?

i) Qual a proporção de árvores com diâmetro entre 10 e 20cm?

j) Se 25% das menores árvores forem cortadas, qual o diâmetro mínimo das árvores remanescentes?

k) Qual o diâmetro mínimo para uma árvore estar entre as 5% maiores árvores?

l) Se 35% das menores árvores forem cortadas, qual o diâmetro mínimo das árvores remanescentes?

m) Qual o diâmetro mínimo para uma árvore estar entre as 1% maiores árvores?

Conceitos-Chave

VARIÁVEL ALEATÓRIA CONTÍNUA - FUNÇÃO DE DENSIDADE - FUNÇÃO DE DENSIDADE NORMAL - PROBABILIDADE NA DISTRIBUIÇÃO NORMAL - PROPRIEDADES DA DISTRIBUIÇÃO NORMAL - PADRONIZAÇÃO - DISTRIBUIÇÃO NORMAL PADRONIZADA - CÁLCULO DE PROBABILIDADES NA NORMAL PADRONIZADA - CÁLCULO DE PROBABILIDADES COM A DISTRIBUIÇÃO NORMAL

Leitura

[WALLIS & ROBERTS, 1964] volume 2, p.470-498. **[ROCHA, 1975]** p.234-241

Exercícios

11.1 As árvores de uma floresta de *Eucalyptus grandis* têm altura com distribuição Normal de média 27m e desvio padrão 3m. Pergunta-se:

a) Esquematize o gráfico da distribuição.

Qual a proporção das árvores com altura:

b) maior que 30m; c) menor que 20m; d) maior que 25m; e) menor que 35m; f) entre 20 e 30m; g) entre 24 e 30m; h) entre 20 e 25m; i) entre 30 e 35m.

Qual a altura mínima da floresta se forem cortadas

j) 10% menores árvores; k) 30% menores árvores; l) 40% menores árvores; m) 50% menores árvores.

Qual a altura que uma árvore precisa ter para estar entre

n) as 10% mais altas; o) as 5% mais altas; p) as 1% mais altas.

11.2 Uma floresta de Pinus taeda tem o DAP (diâmetro a altura do peito) que segue distribuição Normal com média de 21cm e desvio padrão de 5cm.

a) Se todas as árvores com DAP < 18cm forem cortadas qual é a proporção de árvores cortadas?
b) Se um melhorista florestal selecionar as 2.5% maiores árvores da floresta, qual o DAP mínimo das árvores selecionadas?
c) Um Engenheiro Florestal deseja cortar 20% das árvores a partir das menores árvores. Qual é o DAP máximo das árvores a serem cortadas?
d) Uma fábrica de laminação só processa árvores com DAP entre 18 e 27cm. Qual a proporção de árvores desta floresta que poderiam ser utilizadas na fábrica?

11.3 A precipitação anual numa dada localidade segue distribuição Normal com média de 1200mm e desvio padrão de 150mm. Pergunta-se:

a) Qual a precipitação para um ano que você consideraria **extremamente chuvoso**? Explique.
b) Qual a precipitação para um ano que você consideraria **extremamente seco**? Explique.

11.4 Y é uma variável aleatória contínua com distribuição Normal e média 20 e variância 4. Calcule:
a) $P(18 < Y < 23)$ b) $P(12 < Y < 19)$ c) $P(21 < Y < 23)$ d) $P(Y < 21 \mid Y > 18)$ e) $P(Y > 19 \mid Y < 23)$

11.5 X é uma variável aleatória com distribuição Normal de média 100 e variância 36. Calcule:
a) $P(X < 95)$ b) $P(X > 110)$ c) $P(90 < X < 108)$ d) $P(90 \le X \le 108)$
e) $P(X > 92 \mid X < 110)$ f) $P(X < 100 \mid X > 90)$

12. Contornando as Limitações dos Dados: "Distribuições Amostrais II"

Distribuição Amostral da Média

Se amostras aleatórias de tamanho n forem tomadas de uma população com média ***m*** e desvio padrão s, então a distribuição amostral de $\overline{X}_n$ tem as seguintes propriedades:

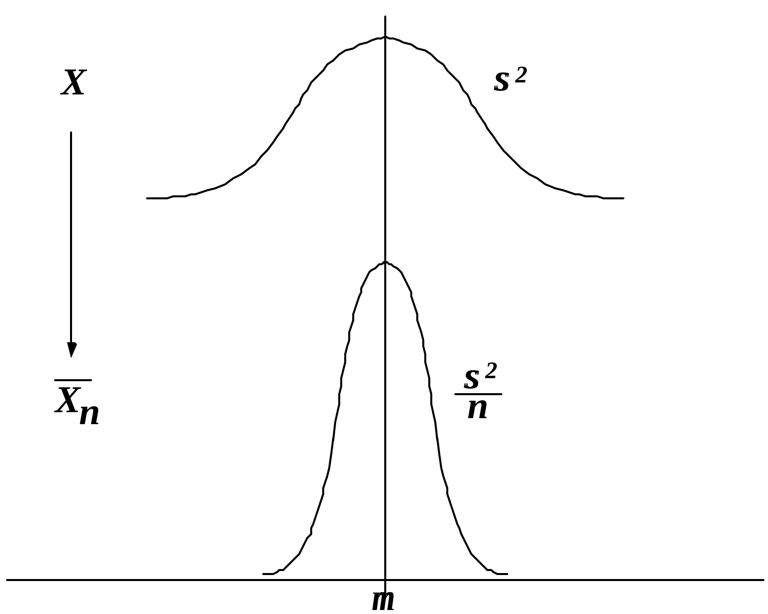
1. $m_{\overline{X}} = m \Rightarrow$ $\overline{X}_n$ é um estimador sem viés de ***m*.**

2. $s^2_{\overline{X}} = \dfrac{s}{n} \Rightarrow s_{\overline{X}} = \dfrac{s}{\sqrt{n}}$

Se o tamanho da amostra cresce, o desvio padrão da média amostral decresce.

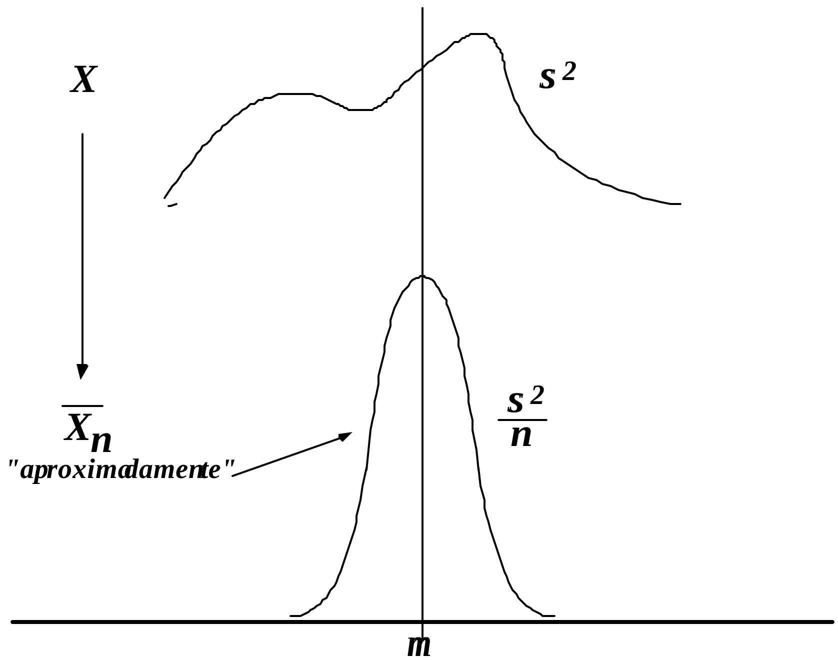
3. Se a população original tem distribuição Normal, então $\overline{X}_n$ também tem distribuição Normal:

$$\overline{X}_n \approx Normal(m, \frac{s^2}{n})$$



4. Teorema Central do Limite:
Se a população original tem uma distribuição qualquer, para n "SUFICIENTEMENTE GRANDE" $\overline{X}_n$ será "APROXIMADAMENTE" Normal:

$$\overline{X}_n \rightarrow aproximadamente\ \ Normal(m, \frac{s^2}{n})$$



Exercício 12.A: Esquema da Distribuição Amostral da Média

Esquematize a distribuição de X e de $\overline{X}$, assumindo que *X* tem distribuição Normal com média 30, desvio padrão 25 e que o tamanho de amostra foi 25.

Exercício 12.B: Produção de Resina em Pinus elliottii

Seja *X* a produção anual de resina de árvores de Pinus elliottii. Suponha que *X* tem distribuição Normal com média 2.3 kg e desvio padrão 0.7kg.

a) Esquematize a distribuição de *X*.

b) A proporção de árvores que produz mais do que 2.8kg será maior ou menor do que 0.5? Por que ?

c) Calcule a proporção de árvores que produzirá mais do que 2.8 kg.

d) Foi realizado uma amostra aleatória de 16 árvores. É mais provável ou menos provável (em relaçao ao item (b)) que a produção média das 16 árvores amostradas seja maior do que 2.8 kg? Desenhe uma figura para embasar a sua resposta.

e) Qual a probabilidade de que a produção média das 16 árvores amostradas seja maior que 2.8 kg?

f) Uma amostra aleatória de 49 árvores foi tomada. É mais provável ou menos provável (em relaçao ao item (d)) que a produção média das 16 árvores amostradas seja maior do que 2.8 kg? Desenhe uma figura para embasar a sua resposta.

g) Qual a probabilidade de que a produção média das 49 árvores amostradas seja maior que 2.8 kg?

Distribuição Amostral de Proporções

p é a proporção das unidades de uma população que possuem uma certa característica (proporção de "sucessos"), como por exemplo:

- árvores com cancro numa floresta de eucalipto;
- árvores cobertas por cipós numa mata;
- capivaras do sexo masculino.

Se amostras aleatórias de tamanho n forem tomadas de uma população com proporçào p, então a distribuição amostral de $\hat{p}_n$ tem as seguintes propriedades:

1. $m_{\hat{p}} = p \Rightarrow$ $\hat{p}_n$ é um estimador sem viés de p.

2. $$s_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Se o tamanho da amostra cresce, o desvio padrão da proporçao amostral decresce.

3. Se a população original tem uma distribuição qualquer, para n "SUFICIENTEMENTE GRANDE", $\hat{p}_n$ será "APROXIMADAMENTE NORMAL":

$$\hat{p}_n \rightarrow aproximadamente \ \ Normal(p, \frac{p(1-p)}{n})$$

Exercício 12.C: Visitantes num Parque

Uma proporção de 37% dos visitantes de um parque favorecem a cobrança de taxas de entrada. Uma amostra aleatória de 200 visitantes foi tomada.

a) Qual é o parâmetro? Qual é a estatística?

b) Qual a probabilidade que na amostra de 200 visitantes pelo menos 40% favoreçam a cobrança de taxas?

c) Qual a probabilidade que na amostra de 200 visitantes, a proporção dos que favorecem a cobrança de taxas fique entre 35 e 39% ?

d) Uma nova amostra de 10 visitantes foi tomada. Qual a probabilidade de que pelo menos 50% dos visitantes na amostra favoreçam a cobrança de taxas ? É válido utilizar o mesmo método utilizado em (b) e (c) ? Por que? Qual método deveria ser utilizado neste caso para encontrar a probabilidade?

Conceitos-Chave

DISTRIBUIÇÃO AMOSTRAL DA MÉDIA - TEOREMA CENTRAL DO LIMITE - DESVIO PADRÃO DA MÉDIA AMOSTRAL - SUFICIENTEMENTE GRANDE - APROXIMADAMENTE NORMAL - DISTRIBUIÇÃO AMOSTRAL DE PROPORÇÕES - DESVIO PADRÃO DE UMA PROPORÇÃO

Exercícios

12.1 Seja X o número de árvores de cedro por hectare numa região de floresta tropical. Suponha que X tem média igual a 2.2 e desvio padrão igual a 1.4. Seja $\overline{X}$ o número médio de árvores de cedro por hectare numa reserva de 100 ha (n=100).

a) Qual a distribuição de $\overline{X}$? Desenhe um gráfico e explique.
b) Qual a probabilidade do número médio de árvores de cedro por hectare na reserva seja menor do que 2?
c) Qual a probabilidade do número total de árvores de cedro na reserva ser 195?

12.2 Numa floresta de *Eucalyptus saligna*, a proporção de árvores mortas é de 4%. Foi tomada uma amostra aleatória de 100 árvores.

a) Qual a distribuição aproximada da proporção de árvores mortas na amostra de 100 árvores? Desenhe uma figura e explique.
b) Qual a probabilidade de que no máximo 7% das árvores na amostra estejam mortas?
c) Qual a probabilidade de que a proporção de árvores mortas na amostra fique entre 2 e 6%.
d) Qual o intervalo centrado na média que compreende a proporção de árvores mortas na amostra com uma probabilidade de 95%?
e) Uma nova amostra aleatória de 8 árvores foi tomada. Qual a probabilidade de que haja no máximo uma árvore morta nesta amostra? Pode ser utilizado o mesmo método utilizados nos itens anteriores? Por que ?

13. Tomando Decisões: "Julgando Hipóteses"

Introdução

A tomada de decisão em Bioestatística envolve o teste de HIPÓTESES. Exitem 2 tipos de hipóteses complementares: HIPÓTESE NULA (H_0) e HIPÓTESE ALTERNATIVA (H_a). Ao escolhermos a hipótese nula podemos incorrer num erro TIPO II (aceitar uma hipótese nula falsa), enquanto ao escolhermos a hipótese alternativa podemos fazer um erro TIPO I (rejeitar uma hipótese nula verdadeira). A possibilidade de cometermos um erro tipo I é chamada de valor-p e dizemos que um teste é dito SIGNIFICATIVO, isto é, rejeita-se a hipótese nula, quando esse valor é inferior ao NÍVEL DE SIGNIFICÂNCIA (**a**). O valor-p sempre depende DO TAMANHO DA AMOSTRA. É importante distinguir um RESULTADO ESTATISTICAMENTE SIGNIFICATIVO de um resultado importante.

Formulando Hipótese Nula e Alternativa: Exemplos

- Hipótese Nula e hipótese Alternativa estabelecem duas possibilidades para uma decisão.
- Essas duas possibilidades devem ser mutuamente exclusivas, isto é, ao rejeitarmos H_0 estaremos automaticamente aceitando H_a , e vice-versa.
- Para que H_0 seja testável quantitativamente, ela deve ser estabelecida da forma mais simples e objetiva possível.

EXEMPLOS:

a) Um estudo sugere que tomando um comprimido de aspirina a cada dois dias durante 20 anos pode reduzir o risco de contrair cancer no cólon pela metade. Contudo, os benefícios só começam a aparecer após uma década. O risco de se desenvolver cancer no cólon durante a vida é de 1 para 16.

H_0 : O risco de cancer no cólon ainda é 1 em 16.

H_a : O risco de cancer no cólon é menor que 1 em 16.

b) Os métodos tradicionais de colheita de madeira em florestas tropicais têm grande impacto sobre a estrutura da floresta. Em geral, a retirada de 10 árvore/ha (diâmetro > 50cm) danifica 20% da cobertura da floresta. Métodos alternativos afirmam que é possível uma redução sensível nesse impacto quanto a localização das árvores a serem abatidas é mapeada e as trilhas de entrada e saida das máquinas é planejada.

H_0 : A colheita danifica 20% da cobertura da floresta.

H_a : O dano da colheita é menor que 20% da cobertura da floresta.

c) O conhecimento popular costuma dizer que o desmatamento provoca seca. Entretanto, estudos indicam que, numa bacia hidrográfica, o corte de áreas florestadas costuma aumentar a quantidade de água nos rios, desde que o impacto do corte sobre o solo seja pequeno.

H_0 : A quantidade de água no rio antes e depois do corte não se altera.

H_a : A quantidade de água no rio após o corte é maior do que antes.

d) Estudos recente indicam que adubação com boro (um micronutriente) aumenta marcantemente a sobrevivência de árvores de *Eucalyptus* durante períodos de seca.

H_0 : A mortalidade de árvores durante a seca é a mesma em florestas adubadas e não adubadas.

H_a : A mortalidade de árvores durante a seca é a menor em florestas adubadas.

Que Erros podemos cometer ?

Existem dois tipos de erro:

Rejeitarmos uma H_0 verdadeira (ou aceitarmos uma H_a falsa) $\longrightarrow$ Erro Tipo I

Não rejeitarmos uma H_0 falsa (ou rejeitarmos uma H_a verdadeira) $\longrightarrow$ Erro Tipo II

A VERDADEIRA HIPÓTESE

		Nula H_0	Alternativa H_a
A SUA ESCOLHA	Nula H_0	Sem erro	**Erro Tipo II**
	Alternativa H_a	**Erro Tipo I**	Sem erro

EXEMPLO:

Hoje à noite você vai a uma festa. A previsão do tempo diz que há 80% de possibilidade de chuva. Você leva um guarda-chuva?

H_0 : Vai chover hoje à noite.

H_a : Não vai chover hoje à noite.

<u>Erro Tipo I</u>: Você rejeita H_0 e, portanto, acredita que não vai chover.
Sai sem o guarda-chuva e se molha !

Erro Tipo II: Você não rejeita $\boldsymbol{H_0}$ e, portanto, aceita que vai chover.
Passa a noite carregando um guarda-chuva sem usa-lo.

Como se Decide Qual Hipótese Aceitar ?

Uma serraria vem utilizando um método tradicional de processar as toras. O novo Engenheiro Florestal deseja implantar e um novo método seria mais eficiente, que reduziria as sobras de madeira (perdas).

$\boldsymbol{H_0}$: O novo método é tão eficiente como o tradicional.

$\boldsymbol{H_a}$: O novo método é mais eficiente que o tradicional.

Questões:
1) Se as perdas de madeira fosse exatamente iguais entre os dois métodos você implantaria o método novo ?

2) Se as perdas no método novo fosse 5% menores que no método tradicional, qual hipótese você aceitaria ?

3) Qual hipótese você aceitaria se a diferença fosse 20% a favor do novo método ?

4) Qual o tamanho da diferença entre o método tradicional e novo método que lhe daria segurança suficiente para rejeitar a hipótese nula ?

- Rejeitar a Hipótese Nula significa, geralmente, mudar um procedimento tradicional em favor de um novo procedimento.
- Esta mudança sempre acarreta custos e evita-se mudar quando não se tem certeza que o novo procedimento será de fato melhor que o tradicional.
- Assim, na maioria das pesquisas estamos mais preocupados em controlar o Erro Tipo I.
- Em estatística, se utiliza a seguinte terminologia para designar a probabilidade de se cometer um erro no teste de hipóteses:

α - é a probabilidade de se cometer um Erro Tipo I (nível de significância),
é a chance de se rejeitar $\boldsymbol{H_0}$ quando ela é verdadeira
(se aceitar uma $\boldsymbol{H_a}$ falsa).

β - é a probabilidade de se cometer um Erro Tipo II,
é a chance de não se rejeitar $\boldsymbol{H_0}$ quando ela é falsa
(se rejeita uma $\boldsymbol{H_a}$ verdadeira).

Exemplo 13.A: Floresta Plantada ou Floresta Nativa ?

A SITUAÇÃO

No final do século passado, Manuel Gomes Acher foi incumbido por D. Pedro II de reflorestar as encostas dos morros na região na Tijuca (Rio de Janeiro). As repedidas secas que a cidade do Rio de Janeiro vinha sofrendo, com a conseqüente falta, de água era atribuida ao desmatamento dos morros. Acher cumpriu sua missão com bastante eficiência e um visitante passeando hoje pelo Parque Nacional da Tijuca terá dificuldade em saber se a floresta que observa é nativa ou foi plantada pelo "major" Acher.

HIPÓTESES

Um pesquisador deseja iniciar um projeto sobre o impacto urbano da cidade do Rio de Janeiro sobre a Floresta da Tijuca. O projeto, entretanto, deve ser instalado em área de floresta nativa. O pesquisador possue como informação inicial os gráficos abaixo que apresentam a frequência de árvores observadas em diferentes classes de tamanhos, em áreas plantadas e áreas nativas. Na escolha da área apropriada para o projeto, o pesquisador trabalha com a seguintes hipóteses:

H_0 : A área escolhida é de floresta nativa.
H_a : A área escolhida é de floresta plantada.

A partir de mapas o pesquisador selecionou uma certa área. Quais os erros que podem ocorrer:

Erro tipo I = Rejeita H_0 quando H_0 é verdadeira
= Conclue que a área é plantada, quando de fato a área é nativa
= Procura uma nova área, embora a área selecionada seja nativa.

Erro tipo II = Aceita H_0 quando H_a é verdadeira (Não rejeita H_0 quando H_0 é falsa)
= Conclue que a área é nativa, quando de fato a área é plantada
= Estabelece o projeto na área errada.

Regra de decisão

Selecionada a área, o pesquisador sortea uma árvore e a mede (amostra de tamanho n=1).

Suponha que a árvore sorteada tenha diâmetro de 80 cm. Você concluiria que a área é nativa ou plantada ? Por que ? E se a árvore sorteada tivesse 20 cm ?

REGRA DE DECISÃO:
Regra que estabelece, com base em dados obtidos, quando a hipótese nula (H_0) é rejeitada.

Várias regras de decisão diferente são possíveis e os níveis de **a** e **b** dependem de qual regra de decisão é utilizada. Verificando a chance de se sortear árvores de diferentes tamanhos em ambas áreas de floresta podemos verificar isso:

Tamanho da Árvore Sorteada (diâmetro em cm)	Chance da árvore ser da área:	
	Nativa	Plantada
20	22 / 40	3 / 40
40	10 / 40	7 / 40
60	4 / 40	12 / 40
80	2 / 40	11 / 40
100	2 / 40	5 / 40
120	0 / 40	2 / 40
Total	40 / 40	40 / 40

<u>Exercício</u>:

Suponha que a H_0 fosse rejeitada se a árvore sorteada tivesse diâmetro de 60 cm ou mais. Estabeleça os níveis α e β.

Rejeita H_0 se o diâmetro ≥ 60 cm $\Rightarrow$ $\alpha = 8 / 40 = 0.20$
$\Rightarrow$ $\beta = 10 / 40 = 0.25$

<u>Exercício</u>:

O que aconteceria com α e β se a regra de decisão estabelece um diâmetro limite maior para rejeitar H_0 ?

Rejeita H_0 se o diâmetro ≥ 100 cm $\Rightarrow$ $\alpha = 2 / 40 = 0.05$
$\Rightarrow$ $\beta = 33 / 40 = 0.83$

<u>Exercício</u>:

No exemplo acima, é possível se manter simultaneamente α e β pequenos ?

Não. Eles estão relacionados de modo inverso. Ao se diminuir a chance de um erro tipo I (α) se aumenta a chance de um erro tipo II (β). A única maneira é trabalhar com o tamanho da amostra.

Nível de Significância

Para que as decisões tomadas por diferentes pessoas tenha uma certa coerrência entre si, cada ramo de atividade costuma estabelecer níveis mínimos para a ocorrência do erro tipo I (α).

Esse nível é chamado de nível de significância . Na área florestal, as decisões devem ser tomadas com um *nível de significância* (α) de pelo menos 0.05 (5%). Um teste de hipótese é chamado *significativo* quando $\boldsymbol{H_0}$ é rejeitada por uma regra de decisão que possui α = nível de significância (0.05 na área florestal).

NÍVEL DE SIGNIFICÂNCIA:
É o nível de a (chance de se cometer um erro tipo I) associado com uma regra de decisão.

<u>Exercício</u>:

Qual o nível de significância para as seguintes regras de decisão:

a) Rejeita Rejeita $\boldsymbol{H_0}$ se o diâmetro ≥ 40 cm.

b) Rejeita Rejeita $\boldsymbol{H_0}$ se o diâmetro ≥ 60 cm.

c) Rejeita Rejeita $\boldsymbol{H_0}$ se o diâmetro ≥ 100 cm.

Qual das regras acima deveria ser utilizada, considerando que a hipótese sendo testada é da área de ciências florestais ?

Valor-p

Digamos que a regra de decisão estabelecida tem nível de significância de 10% (α = 0.10), isto é, a regra é:

Rejeita $\boldsymbol{H_0}$ se o diâmetro da árvore selecionada for ≥ 80 cm.

Caso 1: a árvore selecionada tem diâmetro de 60 cm.

Nesse caso, $\boldsymbol{H_0}$ não é rejeitada, pois 60 cm < 80 cm.

Entretanto, se ignorarmos a regra estabelecida, e rejeitarmos $\boldsymbol{H_0}$ as chance de cometermos um erro tipo I será α = 8 / 40 = 0.20.

Caso 2: a árvore selecionada tem diâmetro de 100 cm.

Nesse caso, $\boldsymbol{H_0}$ é rejeitada, pois 100 cm > 80 cm.

As chance de cometermos um erro tipo I será α = 2 / 40 = 0.05.

VALOR-P:
Valor de a associado aos dados observados, isto é, chance do erro tipo I caso se $\boldsymbol{H_0}$ rejeitada com base nos dados observados.

Conclusão:

Valor-p ≤ Nível de significância ⇒ H_0 é rejeitada

Valor-p > Nível de significância ⇒ H_0 não é rejeitada

Tamanho de Amostra e Chance de erros

Suponhamos agora que o pesquisador decidiu aumentar o tamanho da amostra e sorteou duas árvores. Ele decidiu formular uma regra de decisão com base no diâmetro médio das duas árvores sorteadas. Nesse caso as chances que envolvem duas árvores.

Por exemplo, a chance de selecionarmos duas árvores com diâmetro de 20 cm é:
área nativa = 22 x 22 = 484; área plantada = 3 x 3 = 9.

OBS.: Se lembrarmos que cada número na verdade conta como 1.000.000 ($= 10^6$) árvores, os cálculos exatos seriam: área nativa = (22.000.000 x 21.999.999) / (10^{12} x 10^{12}) = $4.84 \times 10^{14} / 10^{12} \approx 484$;
área plantada = (3.000.000 x 2.999.999) / (10^{12} x 10^{12}) = $9.00 \times 10^{12} / 10^{12} \approx 9$.

Portanto, podemos ignorar o fato que não estaremos selecionando a mesma árvore duas vêzes.

Diâmetros das Duas Árvores Póssiveis de Serem Selecionadas		Valor Médio dos Diâmetros	Número de Possibilidades de se selecionar 2 árvores	
			Nativa	Plantada
20	20	20	484	9
20	40	30	220	21
20	60	40	88	36
20	80	50	44	33
20	100	60	44	15
20	120	70	0	6
40	40	40	100	49
40	60	50	40	84
40	80	60	20	77
40	100	70	20	35
40	120	80	0	14
60	60	60	16	144
60	80	70	8	132
60	100	80	8	60
60	120	90	0	24
80	80	80	4	121
80	100	90	4	55
80	120	100	0	22
100	100	100	4	25
100	120	110	0	10
120	120	120	0	4
		Número Total	1104	976

Sumarizando os resultados em termos apenas de valor médio dos dois diâmetros temos:

Valor Médio dos Diâmetros (cm)	Número de Possibilidades de se selecionar 2 árvores	
	Nativa	Plantada
20	484	9
30	220	21
40	188	85
50	84	117
60	80	236
70	28	173
80	12	195
90	4	79
100	4	47
110	0	10
120	0	4
Número Total	1104	976

<u>Exercício</u>:

Utilizando a tabela acima, resolva os exercícios:

1) Qual o nível de significância para as regras de decisão:

a) Rejeita H_0 se a média dos diâmetros for $\geq$ 100 cm.

b) Rejeita H_0 se a média dos diâmetros for $\geq$ 80 cm.

c) Rejeita H_0 se a média dos diâmetros for $\geq$ 60 cm.

2) Qual o nível de β associado a cada regra de decisão no exercício anterior.
O que acontece com α à medida que o limite para rejeição diminui ?

3) Suponha que as duas árvores selecionadas tiveram média de diâmetro igual a 70cm.

a) H_0 é rejeitada ao nível de 5% de significância?

b) Qual o valor-p ?

4) Para o nível de significância de 5%:

a) Estabeleça a regra de decisão quando quando o tamanho da amostra é de uma árvore.

b) Estabeleça a regra de decisão quando quando o tamanho da amostra é de duas árvore.

c) O que acontece com β (mantendo-se α = 0.05 fixo) quando o tamanho da amostra (n) é aumentado ?

```
Área Nativa (Hipótese Nula)
Número      |
de Árvores  |      X
(1.000.000) |      X
     20     |      X
            |      X
            |      X
            |      X
            |      X
     15     |      X
            |      X
            |      X
            |      X                                 α
            |      X
     10     |      X     X
            |      X     X
            |      X     X
            |      X     X
            |      X     X
      5     |      X     X
            |      X     X     X
            |      X     X     X
            |      X     X     X     X     X
            |      X     X     X     X     X
            ------------------------------------------
                  20    40    60    80    100   120
                Centro de Classes de Tamanho  (diâmetro em cm)

Área Plantada (Hipótese Alternativa)
Número de Árvores (1.000.000)
         20 |
            |
            |
            |      β
            |
         15 |
            |
            |
            |                  X
            |                  X     X
         10 |                  X     X
            |                  X     X
            |                  X     X
            |            X     X     X
            |            X     X     X
          5 |            X     X     X     X
            |            X     X     X     X
            |      X     X     X     X     X
            |      X     X     X     X     X     X
            |      X     X     X     X     X     X
            -------------------------------------------
                  20    40    60    80    100   120
                Centro de Classes de Tamanho (diâmetro em cm)
```

Conceitos-Chave

HIPÓTESES - HIPÓTESE NULA - HIPÓTESE ALTERNATIVA - ERRO TIPO I - ERRO TIPO II - VALOR-P - NÍVEL DE SIGNIFICÂNCIA - ALFA (α) - TAMANHO DE AMOSTRA - DIFERENÇA SIGNIFICATIVA - TESTE SIGNIFICATIVO - RESULTADO ESTATISTICAMENTE SIGNIFICATIVO - RESULTADO IMPORTANTE

Exercícios

13.1 Nos casos abaixo, defina a hipótese nula (H_0) e a hipótese alternativa (H_a) apropriadas.

a) Alguns médicos afirmam que não é possível tratar resfriados e que o próprio organismo desenvolve defesas para neutralizá-lo. Entretanto, uma companhia farmacêutica deseja lançar no mercado o remédio Resfriatim, que num prazo de 5 a 7 dias *cura* qualquer resfriado. Você foi encarregado de organizar um experimento para testar o Resfriatim.

b) Um dos grandes problemas sociais do Brasil é a concentração de renda. O governo alega que após o plano real houve uma sensível melhora desse problema. Como você testaria as alegações do governo ?

c) Com base em pesquisa de opinião realizada no 19o. Salão Internacional do Automóvel (São Paulo, 1996), a Chevrolet anuncia: “Vectra, o melhor carro”. A pesquisa consultou 439 pessoas e o Vectra foi apontado por 24% dos entrevistados como o melhor carro nacional. Vectra é o melhor carro nacional ?

d) Você é diretor de uma fábrica de chapas de madeira (chapas duras). A companhia deseja duplicar a produção da sua fábrica pois o mercado está em franca expansão. Estudando o processo de produção na sua fábrica você chega a conclusão de que o consumo de madeira crescerá em 1.300.000m^3 de madeira/ano com a duplicação. O gerente florestal afirma que as florestas da companhia são capazes de produzir até 1.320.000m^3 de madeira/ano a mais do que vêm produzindo. Será que a floresta é de fato capaz de produzir madeira suficiente para duplicação?

e) *Eucalyptus grandis* é uma das espécies arbóreas de maior produtividade quando plantada no Estado de São Paulo. Entretanto, um experimento mostrou que quando a floresta é plantada sem preparo de solo e sem adubação inicial, *Eucalyptus cloeziana* pode alcançar produtividades de 15 a 20% maiores que *E. grandis*. Quais as hipóteses de um experimento cujo objetivo é comparar essas duas espécies?

f) Sabe-se que numa floresta tropical não perturbada a abundância de espécies “pioneiras” fica em torno de 10%. Com aumento de perturbações antrópicas a abundância dessas espécies tende a crescer. Na demarcação de uma reserva florestal com área total de 50.000 ha, ecologistas e engenheiros florestais discutem a incorporação de uma área de 7.500 ha onde o levantamento de campo revelou uma abundância de 15% de espécies pioneiras. Será que a área de 7.500 ha foi perturbada ?

13.2 Considere as hipóteses nulas abaixo e escreva as hipóteses alternativas apropriadas. Explique.

a) H_0 : A diversidade de espécies arbóreas em floresta tipo cerradão é igual à diversidade em matas ciliares.

b) H_0 : A taxa de incidência de cancro numa floresta plantada de *E. grandis* é de 3%.
[Você deseja verificar se está ocorrendo um surto de cancro na floresta.]

c) H_0 : Em termos de infraestrutura em parques nacionais, as expectativas de visitantes que chegam de ônibus são iguais às expectativas de visitantes que chegam com veículo próprio.

d) H_0 : O novo processo de branqueamento da pasta de celulose produz um mesmo nível de branqueamento que o processo tradicional.
[Você deseja provar que o novo processo é melhor.]

e) H_0 : O instrumento a laser para medição de árvores resulta no mesmo tempo de medição dos instrumentos ópticos.
[Você deseja mostrar que o instrumento a laser é melhor.]

13.3 O princípio da justiça é que todo acusado é inocente até que seja provado culpado.
a) Qual é a hipótese nula e a hipótese alternativa num julgamento criminal ?
b) Quais são as consequências equivalentes aos erros tipo I e II ?

13.4 Uma empresa florestal deseja introduzir um novo clone de *E. saligna* nas suas plantações comerciais. Foi realizado um experimento comparando o novo clone com o clone tradicionalmente utilizado. As hipóteses nesse experimento foram:

H_0 : A produtividade do novo clone e do clone tradicional são iguais.

H_a : A produtividade do novo clone é maior do que a do clone tradicional.

a) Descreva as conseqüências do erro tipo I.
b) Descreva as conseqüências do erro tipo II.
c) O experimento resultou numa diferença estatísticamente significativa entre os clones. Qual é a hipótese rejeitada ? Por que ?

13.5 Uma serraria que fornece peças de madeira para diversas fábricas de móveis recebeu várias reclamações de que as peças de largura de 10 cm têm apresentado em média uma largura menor do que 9 cm. Você é responsável por verificar essas reclamações a partir de dados da linha de produção.

a) Quais são as hipóteses nula e alternativa ?
b) Qual é o erro tipo I e as suas consequências ?
c) Quais as consequências do erro tipo II ?
d) Suponha que os dados coletados resultaram num teste estatisticamente significativo. Qual a hipótese que seria rejeitada ?

13.6 Numa floresta plantada, foi realizado um levantamento para se verificar a existência de um surto de cancro. A taxa normal de ocorrência de cancro é de 2%, assim as hipóteses testadas foram:

H_0 : A taxa de cancro na floresta é de 2%.

H_a : A taxa de cancro na floresta é maior do que 2%.

a) Quais os erros tipo I e II ?
b) Quais as consequências de ambos tipos de erro ?
c) Se o teste de hipótese **não** for estatisticamente significativo, qual hipótese será rejeitada ?

13.7 O IBAMA é o orgão federal responsável pelo controle de desmatamento. Um proprietário só pode ser multado por desmatamento ilegal se esse for comprovado em campo. Na região amazônica, entretanto, o controle por visitas a propriedades é praticamente impossível. Com a ajuda da NASA, o IBAMA tem utilizado uma areonave de alta altitude equipada com radar. Essa aeronave indica áreas **prováveis** de terem sofrido desmatamento não autorizado e uma equipe de terra do IBAMA visita a propriedade em questão. Caso o desmatamento seja comprovado na visita, o proprietário é multado. Suponha que uma dada área foi indicada como desmatamento ilegal provável.

a) Defina as hipóteses $\boldsymbol{H_0}$ e $\boldsymbol{H_a}$.
b) Defina as consequências do erro tipo I.
c) Defina as consequências do erro tipo II.
d) Quais consequências são mais piores ?

13.8 O número de peças defeituosas produzida por uma destopadeira segue os gráficos apresentados abaixo:

```
           REGULADA                              DESREGULADA
 x                                                               x
 x    x                                                     x    x
 x    x                                                     x    x
 x    x                                                     x    x
 x    x                                                     x    x
 x    x    x    x                                 x    x    x    x
 x    x    x    x    x    x             x    x    x    x    x    x
-------------------------------      -----------------------------
 1    2    3    4    5    6             1    2    3    4    5    6
    peças defeituosas/hora                 peças defeituosas/hora
```

As hipóteses a serem testadas são: $\boldsymbol{H_0}$: A destopadeira está regulada.
$\boldsymbol{H_a}$: A destopadeira não está regulada.

a) Utilizando um nível de significância de 5% (**a**), qual hipótese seria rejeitada se a regra de decisão fosse: Rejeitar $\boldsymbol{H_0}$ caso ocorra 6 ou mais peças defeituosas numa hora.

b) Qual a probabilidade de um erro tipo II (**b**) para a regra de decisão acima ?

c) Utilizando um nível de significância de 10% (**a**), qual hipótese seria rejeitada se a regra de decisão fosse: Rejeitar $\boldsymbol{H_0}$ caso ocorra 4 ou mais peças defeituosas numa hora.

d) Qual a probabilidade de um erro tipo II (**b**) para a regra de decisão acima ?

e) Quais as consequências dos erros tipo I e II ?

f) Qual seria o valor-p para a seguinte regra de decisão: Rejeitar $\boldsymbol{H_0}$ caso ocorra 5 ou mais peças defeituosas numa hora? $\boldsymbol{H_0}$ seira rejeitada num nível de significância de 10%? Por que ?

g) Qual seria o valor-p para a seguinte regra de decisão: Rejeitar $\boldsymbol{H_0}$ caso ocorra 3 ou mais peças defeituosas numa hora? $\boldsymbol{H_0}$ seira rejeitada num nível de significância de 10%? Por que ?

13.9 Impactos ambientais em florestas tropicais resultam, em geral, numa redução no número de espécies arbóreas. A tabela abaixo apresenta a freqüência com que se observar um dado número de espécies em parcelas de 1 hectare em florestas maduras e florestas degradadas. A questão básica é detectar se uma dada área florestal foi degradada, portanto, as hipóteses sendo testadas são:

H_0 : Não houve degradadação da floresta.

H_a : Houve degradação da floresta.

Número de Espécies Arbórea em parcelas de 1 hectare	Floresta Madura	Floresta Degradada
40	1	25
50	2	20
60	4	17
70	8	13
80	10	10
90	13	8
100	17	4
110	20	2
120	25	1
Total de parcelas	100	100

Parte I: Apenas uma parcela foi observada.

A regra de decisão é Rejeitar H_0 se o número de espécies na parcela for 50 ou menos.

a) Qual o valor-p dessa regra? H_0 seria rejeita para **a** = 0.05 ?
b) Qual é o valor de **b** correspondente a regra de decisão acima?
c) Qual a regra de decisão que garante nível de significância de 1% ?

Parte II: Foram observadas 2 parcelas.

A regra de decisão é Rejeitar H_0 se o **número mínimo** de espécies por parcela for 50 ou menos.

d) Qual o valor-p dessa regra? H_0 seria rejeita para **a** = 0.05 ?
e) Qual é o valor de **b** correspondente a regra de decisão acima?
f) Qual a regra de decisão que garante nível de significância de 1% ?

13.10 Comente os resultados apresentados abaixo em termos de diferença estatisticamente significante e diferença importante.

a) Na comparação de dois lotes de mudas de árvores encontrou-se uma diferença de 0.1 cm na altura média que foi considerada estatisticamente significativa (**a** =0.05).

b) Num experimento visando detectar o efeito de adubação em florestas plantadas, a hipótese nula não foi rejeitada (**a** =0.05), embora a produtividade média diferisse em 20 m3/ha.ano entre parcelas adubadas e não adubadas.

14. *Testando Hipóteses sobre os Parâmetros de uma População*

Testando Hipóteses Estatísticas

Hipótese Estatísticas sempre surgem
na forma de
DUAS HIPÓTESE COMPLEMENTARES

HIPÓTESE NULA H_0
Geralmente estabelece uma relação
de igualdade entre duas quantias

HIPÓTESE NULA H_α
Estabelece uma relação
de desigualdade entre duas quantias

Ao testarmos
uma hipótese
estatística
podemos
cometer dois
tipos de erro

ERRO TIPO I
Rejeitar uma H_0 verdadeira.
(implica em aceitar uma H_α falsa)

Porb. erro tipo I
α
NÍVEL DE
SIGNIFICÂNCIA

ERRO TIPO II
Não rejeitar uma H_0 falsa.
(implica em refjeitar uma H_α verdadeira)

Porb. erro tipo II
β

Os erros tipo I são,
geralmente, de
consequências mais
danosas

Decisões estatísticas
sempre controlam a
margem de erro tipo I.

Na Engenharia Floresta o
nível de significância é de 5%,
isto é, a probabilidade de
erro tipo I aceitável é
de no máximo 5%
($\alpha = 0.05$)

Exemplo 14.A Mortalidade de Árvores numa Floresta

Após vários anos de acompanhamento de parcelas permanentes, uma Eng. Florestal concluiu que, das árvores que morrem num fragmento florestal, 75% são devido ao abafamento da copa por cipós. No ano seguinte, ocorreu um período prolongado de intensa seca e 30 árvores morreram durante o ano, das quais 24 mortes podem ser atribuidas ao efeito dos cipós. A Eng. Florestal afirma que a proporção de árvores que morreram devido aos cipós (p) foi maior neste ano de seca intensa do que nos anos anteriores.

a. Estabeleça H_0 e H_α.

   H_0: A proporção de árvores mortas por cipós no ano de seca intensa é igual a 75%.

   H_α: A proporção de árvores mortas por cipós no ano de seca intensa é maior que 75%.

b. Estabeleça H_0 e H_α de forma quantitativa.

   $H_0 : p = 0.75$

   $H_\alpha : p > 0.75$

c. De acordo com a afirmação da Eng. Florestal, a hipótese nula foi rejeitada ou aceita? Explique

d. Se o teste de hipótese utilizou o nível de 5% de significância, o valor-p foi maior ou menor que 0.05? Explique.

e. Qual o tamanho de amostra utilizado para os resultados no ano de seca intensa?

Testando Hipóteses sobre a Proporção de uma População

Como a hipótese acima pode ser testada?

Como a taxa de mortalidade do último ano (ano de seca intensa) foi baseada numa amostra, o valor

$$\frac{24}{30} \times 100 = 80\%$$

é uma estatística da amostra e deve ser representado por $\widehat{p}$. Nós sabemos que $\widehat{p}$ varia de amostra para amostra. Se a amostragem fosse repedita e outras 30 árvores mortas observadas, nós provavelmente obteríamos um outro valor para $\widehat{p}$.

Como já foi visto, a distribuição amostral de $\widehat{p}$ tem as seguintes propriedades:

1. a média da distribuição de $\widehat{p}$ será p;
2. a variância da distribuição de $\widehat{p}$ será
$$\frac{p(1-p)}{n};$$
3. a distribuição de $\widehat{p}$ será aproximadamente Normal.

Com base nestas propriedades sabemos que a variável aleatória:

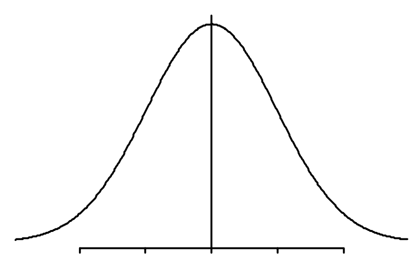
$$Z = \frac{\widehat{p} - p}{\sqrt{p(1-p)/n}}$$

terá distribuição aproximadamente Normal Padronizada, isto é, distribuição Normal com média 0 e desvio padrão 1.

Assumindo que a hipótese nula é verdadeira ($H_0 : p = 0.75$), $\widehat{p}$ seria aproximadamente Normal com média 0.75 e desvio padrão

$$\sqrt{\frac{0.75(1-0.75)}{30}}$$

(distribuição ao lado).
Podemos agora quantificar, sob as condições de H_0, a probabilidade de observarmos um valor de $\widehat{p} = 0.80$ ou mais extremo, .



O que quer dizer um valor mais extremo ?

A hipótese alternativa (H_α) estabelece que $p > 0.75$, portanto a Eng. Florestal acredita que a mortalidade no ano de seca intensa é **maior** do que nos anos normais. Se isso de fato ocorrer, o valor amostral $\widehat{p}$ poderia ser ainda maior do que 0.80.

$$H_\alpha : p > 0.75 \quad \Rightarrow \quad P\left[\widehat{p} \geq 0.80\right]$$
$$H_0 : p = 0.75 \quad \Rightarrow \quad P\left[\widehat{p} \geq 0.80\right] = P\left[Z \geq \frac{0.80 - 0.75}{\sqrt{0.75(1-0.75)/30}}\right] = P\left[Z \geq 0.63\right] = 0.2643$$

Portanto, sob H_0, a probabilidade de se observar um valor de $\widehat{p}$ ou maior é 0.2643. Se rejeitarmos H_0 com base no valor observado de $\widehat{p} = 0.80$, teremos a probabilidade de 0.2643 de cometermos um erro tipo I. Logo, 0.2643 é o valor-p, mas a margem de erro tipo I aceitável nas Ciências Florestais é de 0.05 ($alpha$ - nível de significância). Como o valor-p (0.2643) é bem maior do que o nível de significância ($\alpha = 0.05$) nós não rejeitamos H_0 e podemos concluir que, do ponto de vista estatístico, a afirmação da Eng. Florestal não procede.

Teste Formal de Hipóteses

O exemplo acima representa um teste formal de hipóteses estatística a respeito da proporção p em uma população. Este teste é chamado de **teste Z para proporção envolvendo uma amostra**. Vejamos os elementos básicos deste teste:

- A hipótese sendo testada é a respeito da proporção p de uma população que tem determinado atributo. Mais especificamente:

$$H_0 : p = 0.75 \quad \text{versus} \quad H_\alpha : p > 0.75$$

  Note que as hipóteses se referem a proporção na população e não na amostra.

- A decisão a respeito de p foi baseada na proporção da amostra $\widehat{p}$. Mais precisamente no **valor padronizado** de $\widehat{p}$:

$$Z = \frac{\widehat{p} - p_0}{\sqrt{p_0(1-p_0)/n}}$$

  onde p_0 é o **valor hipotetizado** para p na hipótese nula (H_0).

  Esta quantia padronizada é chamada de *estatística de teste* e sua distribuição sob H_0 é aproximandamente Normal, com média 0 e desvio padrão 1.

- O valor-p para o teste é calculado com base em como a hipótese alternativa (H_α) é expressa.

- **Decisão:** um valor-p menor que o nível de significância implica em rejeitar H_0. No exemplo acima, no entanto, o valor-p foi maior que o nível de significância e H_0 não foi rejeitada, concluindo-se que a afirmação da Eng. Florestal não tem embasamento estatístico.

Exemplo 14.B Rabo-de-Raposa

A proporção de árvores com rabo-de-raposa numa floresta de *Pinus oocarpa* é da ordem de 10%. Após um projeto de melhoramento genético, o melhorista florestal afirmou que a proporção de árvores com rabo-de-raposa tinha sido reduzida sensívelmente na população de árvores melhoradas, pois numa amostra aleatória de 100 árvores, apenas 7 apresentaram o problema.

a. Estabeleça as hipóteses nula e alternativa para testar a afirmação do melhorista.

   H_0: ______________________________

   H_α: ______________________________

b. Estabeleça de forma quantitativa as hipóteses nula e alternativa.

   H_0: ______________ versus H_α: ______________

c. Encontre o valor da estatística Z.

d. Encontre o valor-p correspondente à estatística Z.

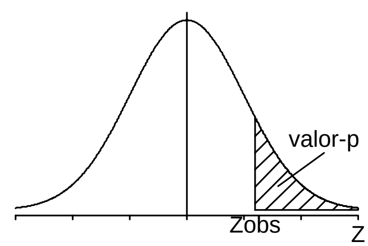
e. Estabeleça a decisão estatística final.

Testes Monocaudais e Testes Bicaudais

A hipótese alternativa define como o valor-p é calculado

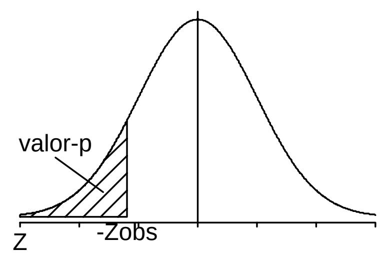
$H_\alpha : p > p_0$

Valor-p: área da cauda a direita do valor observado para estatística de teste sob H_0.



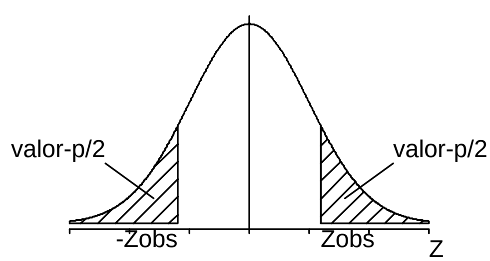
$H_\alpha : p < p_0$

Valor-p: área da cauda a esquerda do valor observado para estatística de teste sob H_0.



$H_\alpha : p \neq p_0$

Valor-p: área nas duas caudas (extremo em ambas direções) fora da região definida pelo do valor observado para estatística de teste sob H_0.
Como a distribuição Normal(0,1) é simétrica, isto implica em dobrar a área de uma das caudas.



Exemplo 14.C A Moeda Justa

O matemático inglês John Kerrick lançou 10000 vêzes uma moeda. Do total de lançamentos, 5064 resultaram em "cara".

a. Existe evidência, ao nível de 5% de probabilidade, de que a moeda não é justa?

Hipóteses:

Dados: $\widehat{p} =$

Estatística: $Z =$

Valor-p:

Decisão:

b. Existe evidência, ao nível de 5% de probabilidade, de que o número de "caras" é maior do que o esperado?

Hipóteses:

Dados: $\widehat{p} =$

Estatística: $Z =$

Valor-p:

Decisão:

Testando Hipóteses sobre a Média de uma População

Até agora foi visto o teste Z para se testar a proporção de uma população. Qual seria o teste mais apropriado para testarmos hipóteses sobre *a média de uma população*?

Exemplo 14.D Testando o pH do Solo

Um cientista deseja saber se o pH de um solo é acido. Ele uma amostra com cinco unidades e obteve os valores de ph:

$$5.8;\ 6.3;\ 6.9;\ 6.2;\ 5.5$$

Considere os seguintes aspectos:

- O cientista considera o solo ácido se o seu pH for menor que 7. Portanto as hipótese a serem testadas são:

  H_0: ______________________________

  H_α: ______________________________

- Utilizando a teoria da distribuição da média amostral podemos utilizar o seguinte raciocínio: Seja X o pH do solo e possui distribuição Normal com média μ e desvio padrão σ, então:

  $\overline{x}_5$ terá média igual μ;

  $\overline{x}_5$ terá desvio padrão igual a $\sigma/\sqrt{5}$;

  $\overline{x}_5$ terá distribuição Normal.

- Logo, pela teoria da distribuição amostral, a estatística:

$$Z = \frac{\overline{x}_5 - \mu}{\sigma/\sqrt{5}}$$

  terá distribuição Normal com média 0 e desvio padrão 1. Poderíamos utilizá-la para testar H_0.

Estas considerações apresentam um problema: a hipótese nula (H_0) nos fornece uma indicação para a média da população (μ), mas não nos dá qualquer dica sobre o desvio padrão da população (σ). O desvio padrão da população (σ) permanece desconhecido e, não modo de se obter a estatística Z com base na amostra tomada pelo cientísta.

Estatística t

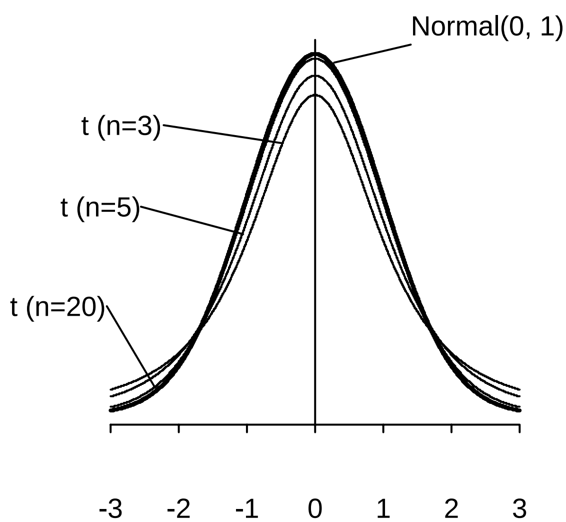
Este tipo de problema já foi estudado no começo do século pelo matemático inglês William Gosset, que usava o pseudônimo "Student" para publicar os seus trabalhos. Student desenvolveu a seguinte estatística:

$$t = \frac{\overline{x}_n - \mu}{s_n/\sqrt{n}}$$

onde s_n é o desvio padrão amostral (amostra de tamanho n).

A estatística t é uma variável aleatória cuja distribuição tem as seguintes propriedades:

1. Tem forma de sino (como a dist. Normal).
2. É simétrica em relação ao zero (como a dist. Normal Padronizada).
3. É mais "achatada" no centro e tem as caudas mais "pesadas".
4. Possui somente um parâmetro: $\nu = n-1$ chamado de graus de liberdade.
5. À medida que $\nu \to \infty$, então $t \to Z$, isto é, para grandes amostras a distribuição de t se aproxima da Normal Padronizada.



É importante notar que para cada valor dos graus de liberdade ($\nu = n - 1$) temos uma distribuição diferente para o t de Student. A distribuição Normal Padronizada é uma única distribuição (o valor dos parâmetros são fixos), mas a distribuição t de Student é na verdade uma *família de distribuições*, pois cada valor particular de ν gera uma distribuição distinta das demais.

Nesta situação, torna-se impossível existir uma tabela que apresente todos os valores de todas as distribuições t possíveis. O Anexo C, no entatno, apresenta os quantis da distribuição t de Student para alguns valores particulares de α (probabilidade na cauda) e uma série de valores distintos de ν.

Exemplo 14.D Testando o pH do Solo

Podemos agora testar as hipótese a respeito do pH do solo utilizando a estatística t. Seja μ o pH médio do solo e $\overline{x}$ a média amostral dos pHs na amostra:

- Hipóteses

  $H_0 : \mu = 7$

  $H_\alpha : \mu < 7$

- Dados

  $\overline{x} = (5.8 + 6.3 + 6.9 + 6.2 + 5.5)/5 = 6.14$

  $s = 0.5320$

- Estatística t:

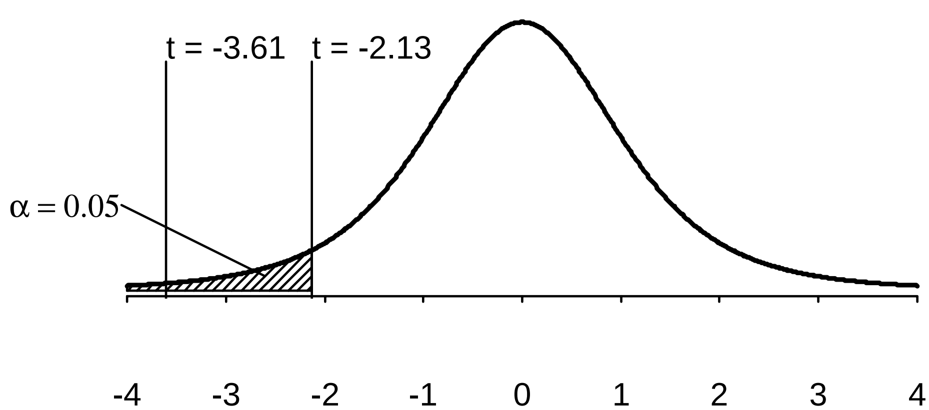
$$t = \frac{6.14 - 7}{0.5320/\sqrt{5}} = -3.61$$

- Valor-p

$$H_\alpha : \mu < 7 \quad \Rightarrow \quad P[\overline{x} < 6.14] \Rightarrow P[t < -3.61] \Rightarrow P[t > 3.61] \quad \text{(Simétrica em relação ao zero)}$$
$$H_0 : \mu = 7 \quad \Rightarrow \quad P[t > 3.61] = \text{valor-p}$$

Na tabela do t de Student (Anexo C) encontramos o valor 2.13 para $t_{\alpha=0.05;\nu=4}$. Logo:

$$\begin{aligned} P[t > 3.61] < P[t > 2.13] \Rightarrow \quad & P[t < -3.61] < P[t < -2.13] \\ \Rightarrow \quad & \text{valor-p} < \alpha = 0.05 \end{aligned}$$



- Decisão: Rejeita-se H_0 (ao nível de 5% de probabilidade) e concluí-se que existe evidência estatística para afirmar que o pH médio do solo é menor que 7.

Teste t para a Média de uma População

O teste estatístico que acabamos de utilizar é chamado de **teste t para a média de uma população** e ele possui as seguintes características:

- A hipótese sendo testada é a respeito da média μ de uma população. No exemplo do pH temos:

$$H_0 : \mu = 7 \quad \text{versus} \quad H_\alpha : \mu < 7$$

  Note que as hipóteses se referem à média da população e não da amostra.

- A decisão a respeito de μ foi baseada na média da amostra $\overline{x}_n$. Mais precisamente na estatística t observada na amostra:

$$t = \frac{\overline{x}_n - \mu_0}{s_n/\sqrt{n}}$$

  onde μ_0 é o **valor hipotetizado** para μ na hipótese nula (H_0).

  Sob a hipótese nula, a estatística t tem distribuição t de Student com parâmetro $\nu = n - 1$ (graus de liberdade.

- O valor-p para o teste é calculado com base em como a hipótese alternativa (H_α) é expressa.

- **Decisão:** um valor-p menor que o nível de significância (α) implica em rejeitar H_0. Foi exatamente o que ocorreu no exemplo do pH. Quando o valor-p é maior que o nível de significância, H_0 não é rejeitada.

Conceitos-Chave

HIPÓTESES ESTATÍSTICAS - HIPÓTESE NULA - HIPÓTESE ALTERNATIVA - ERRO TIPO I - VALOR-P - NÍVEL DE SIGNIFICÂNCIA - PROPORÇÃO NULA POPULAÇÃO - PROPORÇÃO AMOSTRAL - DISTRIBUIÇÃO AMOSTRAL DE UMA PROPORÇÃO - VALOR EXTREMO - TESTE Z PARA PROPORÇÃO - TESTE MONOCAUDAL - TESTE BICAUDAL - MÉDIA DA POPULAÇÃO - MÉDIA AMOSTRAL - DISTRIBUIÇÃO AMOSTRAL DA MÉDIA - ESTATÍSTICA t DE STUDENT - DISTRIBUIÇÃO t DE STUDENT - TESTE t PARA MÉDIA DE UMA POPULAÇÃO.

Exercícios

14.1 Acredita-se que a proporção de tábuas de *Eucalyptus grandis* que apresentam sérios problemas de rachadura de topo após o desdobro é da ordem de 85%. Uma pesquisa com 100 tábuas de um clone de *E. grandis* resultou numa proporção de 78% das tábuas com rachadura de topo. O melhorista florestal garante que este ganho é "significativo", o responsável pela serraria acredita que essa diferença é "devida ao acaso". Qual dos dois tem embasamento estatístico? Explique.

14.2 Um experimento clássico em Percepção Extra Sensorial (PES) consiste em embaralhar um baralho com cartas de cinco naipes (ondas, estrelas, círculos, quadrados e cruzes). À medida que o pesquisador vira uma carta e se concentra nela, a pessoa sendo pesquisada tenta advinhar o naipe da carta. Uma pessoa sem PES tem probabilidade de 1/5 (0.20) de acertar por pura sorte. Já uma pessoa com PES deverá acertar com mais frequência. Isadora acertou 36 em 50 tentativas. Teste ao nível $\alpha = 0.01$ se Isadora tem PES. Encontre o valor-p e estabeleça as suas conclusões.

14.3 Um estudante de estatística usou um programa de computador para testar a hipótese $H_0 : p = 0.50$ contra a hipótese $H_\alpha : p > 0.50$ (teste monocaudal). O programa foi alimentado com 500 observações de uma amostra simples aleatória e o programa retornou o seguinte resultado:

Estatística Observada: $Z = 0.44$ valor-p=0.33

a. Baseado no valor-p acima, o estudante concluiu que existe uma probabilidade de 33% da hipótese nula ser verdadeira. Você concorda ? Explique.

b. Qual seria o valor-p se o teste fosse bicaudal (hipótese alternativa $H_\alpha : p \neq 0.50$)?

14.4 No último Fantesão, o pessoal da Kπxama observou os seguintes dados:

número de pessoas: 521
número de homens: 300
número de cervejas/homem: 4
número de cervejas/mulher: 2.5.

Este ano, espera-se que as taxas de consumo permaneçam as mesmas e que compareçam 700 pessoas. Estimando uma proporção de 60% de homens, o pessoal da Kπxama encomendou 100 caixas de cerveja (cada uma com 24 cervejas). Vai faltar cerveja?

a. Estabeleça as hipóteses apropriadas.

b. Efetue o teste e encontre o valor-p.

c. Com base no teste, qual a sua decisão?

14.5 Um Eng. Florestal deseja saber se a altura média de uma floresta nativa é superior a 20m. Numa amostra de 100 parcelas de inventário, ele obteve média amostral de 23m e desvio padrão amostral de 7.5m. Há evidência estatística (nível de 5% de probabilidade) para se acreditar que a altura média da floresta é superior a 20m?

14.6 Um levantamento do tempo de quebra de motosserras numa empresa florestal resultou numa média de 120 horas e desvio padrão de 30 horas. O Eng. Florestal responsável pela área de colheita deseja saber se ele pode elaborar orçamento anual assumindo que o tempo médio de quebra é de 100 horas.

14.7 Zé do Grude é proprietário de uma fábrica de cola. Como parte do programa de controle de qualidade de sua fábrica, ele amostra toda semana pacotes de 20kg de cola para verificar se o peso dos pacotes não difere muito do peso nominal. Se os pacotes forem mais leves os clientes reclamam, se forem mais pesados a empresa perde dinheiro. Em ambos os casos a máquina de empacotamento teria de ser desligada para investigar o que está ocorrendo.

a. Estabeleça as hipóteses apropriadas.

b. Quais as pressuposições necessárias para testar estas hipóteses?

c. Como um Erro Tipo I afetaria a companhia ?

d. Como um Erro Tipo II afetaria a companhia ?

e. A amostra semanal de 25 pacotes resutou numa média de 19.6kg e num desvio padrão de 0.9kg. Realize o teste e encontre o valor-p.

f. Baseado no resultado do teste, qual seria a sua recomendação?

14.8 São 12:30 hs e você está na fila do RUCALQ, com 33 pessoas na sua frente. Você está ansioso pois tem um compromisso às 13:00 hs. Um aluno de PG do Departamento de Estatística tem acompanhado o tempo de atendimento na fila e obteve os seguintes resultados:

Número de Pessoas	Tempo para Atendê-las (minutos)
35	28
40	30
30	25
39	32
24	20

Será que você estará almoçando até às 13:00 hs?

a. Estabeleça as hipóteses apropriadas.

b. Quais as pressuposições necessárias para testar as hipóteses?

c. Efetue o teste e encontre o valor-p.

d. Com base no teste, qual a sua decisão?

15. Inferência sobre os Parâmetros de uma População: "Intervalo de Confiânça

Teste de Hipóteses

O teste de hipóteses (H_0 versus H_α) envolve:
(1) o parâmetro da população (p.ex.: média μ);
(2) informações de uma amostra
(p.ex.: média $\overline{x}$ e desvio padrão s);
(3) um estatística cujo o comportamento é conhecido
sob H_0 (p.ex.: estatística t).

H_0 é rejeitada se:
valor-p $\leq \alpha$

H_0 não é rejeitada se:
valor-p $> \alpha$

Exemplo 15.A Produção em Florestas de Plantadas

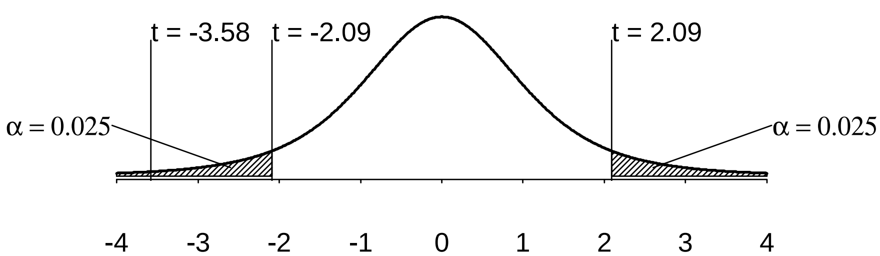
Numa floresta plantada, um conjunto de 20 parcelas de inventário resultou numa produção média de 280 st/ha e desvio padrão de 25 st/ha. O engenheiro florestal deseja saber se a produção média desta floresta é estatisticamente diferente de 300 st/ha.

a. Hipóteses: $H_0 : \mu = 300$ versus $H_\alpha : \mu \neq 300$.

b. Dados: $\overline{x} = 280$ e $s = 25$.

c. Estatística: $t = \frac{280-300}{25/\sqrt{20}} = -3.58$

d. Valor-p:



e. Decisão: Rejeita-se H_0. Há evidências estatísticas de que a produção não é igual a 300 st/ha.

Regra de Decisão: Abordagem Clássica

Como vimos, a hipótese nula (H_0) será sempre rejeitada quando:

$$\text{valor-p} \leq \text{nível de significância } (\alpha).$$

Existe outra forma de se verificar quando a hipótese nula é rejeitada utilizando apenas a informação do nível de significância (α).

Testes Monocaudais

Suponha que no exemplo anterior estívemos testando as hipóteses:

$$H_0 : \mu = 300 \text{ versus } H_\alpha : \mu > 300,$$

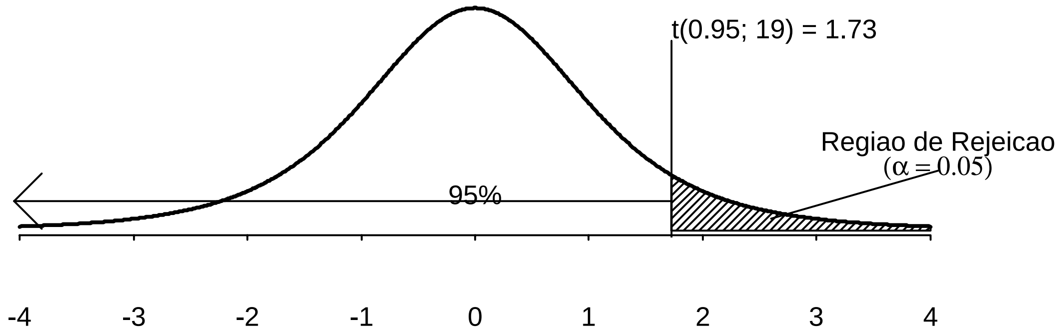
através da mesma estatística t:

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

Rejeita H_0 se $\overline{x}$ for "muito maior" que μ.	→	Rejeita H_0 se o valor de t é "muito grande"	→	Rejeita H_0 se $t \geq c$ c = VALOR CRÍTICO

Qual o valor crítico (c) que corresponde a uma probabilidade de Erro Tipo de I de 5% ?

$$\begin{aligned} \alpha &= P\,[\text{ Erro Tipo I }] \\ &= P\,[\text{ Rejeita } H_0 | H_0 \text{ verdadeira }] \\ &= P\,[t \geq c | t \sim \; t \text{ de Student}(\nu = 19)] = 0.05 \end{aligned}$$



VALOR CRÍTICO $c = t_{0.95;19} = 1.73$	→	REGIÃO DE REJEIÇÃO $\alpha = 0.05$

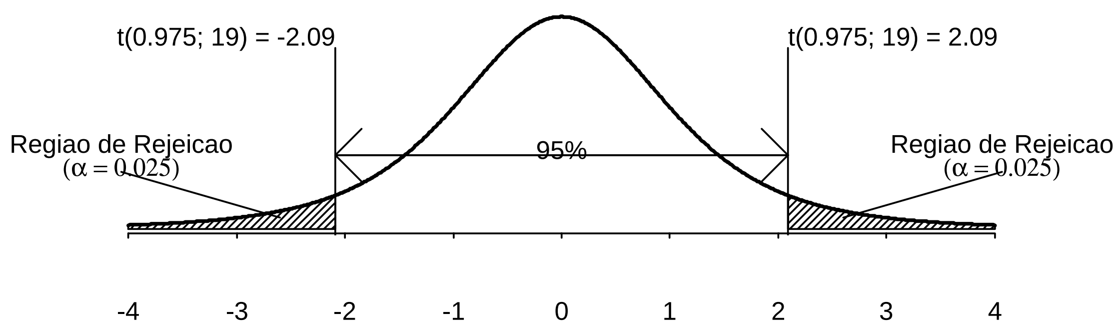
<u>*Testes Bicaudais*</u>

No exemplo de produção temos

Hipóteses Testadas: $H_0 : \mu = 300$, $H_\alpha : \mu \neq 300$ Estatística: $t = [\overline{x} - \mu]/[s/\sqrt{n}]$

Neste caso:

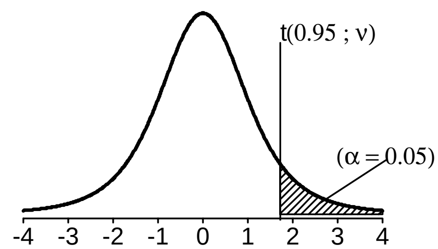
Rejeita H_0 se $\overline{x}$ for mais extremo que μ. ("<u>muito maior</u>" ou "<u>muito menor</u>")	→	Rejeita H_0 se o valor de t é "<u>extremo</u>"	→	Rejeita H_0 se $\lvert t\rvert \geq c \Rightarrow t \geq c$ ou $\lvert t\rvert \geq c \Rightarrow t \leq c$



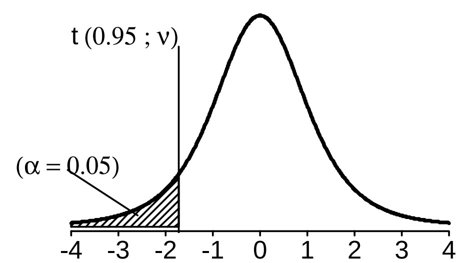
Valor Crítico	Região de Rejeição
$c = t_{0.975;19} = 2.09$	$c = t_{0.975;19} = 2.09 \Rightarrow \alpha = 0.025$ $c = -t_{0.975;19} = -2.09 \Rightarrow \alpha = 0.025$

Região de Rejeição

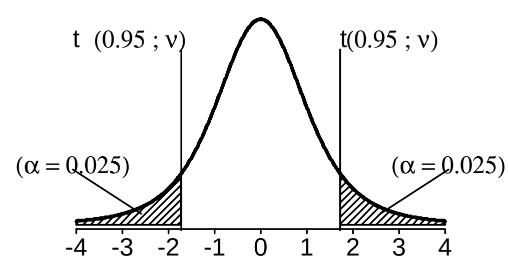
$H_\alpha : \mu > \mu_0$ Rejeita H_0 se $t \geq c$ onde $c = t_{(1-\alpha);\nu}$



$H_\alpha : \mu < \mu_0$ Rejeita H_0 se $t \leq c$ onde $c = t_{(\alpha);\nu}$ (note que $t_{(\alpha);\nu} = -t_{(1-\alpha);\nu}$)



$H_\alpha : \mu \neq \mu_0$ Rejeita H_0 se $t \geq c$ ou $t \leq -c$ onde $c = t_{(1-\alpha/2);\nu}$



Intervalo de Confiança para a Média de uma População

Média amostral ($\overline{x}$) é uma Estimativa Pontual da média da população (μ).	Qual a qualidade desta estimativa pontual ? Podemos confiar nela?

Podemos definir intervalo simétrico ao redor da média amostral ($\overline{x}$) que contém valores "plausíveis" para a média populacional (μ).

Este intervalo pode ser definido de acordo com um dado "nível de confiança".

Utilizamos a estatística t

$$t = \frac{\overline{x} - \mu}{s/\sqrt{n}}$$

para testar as hipóteses

$$H_0 : \overline{x} = \mu \text{ versus } H_\alpha : \overline{x} \neq \mu$$

num determinado nível de significância (α).

Assumindo H_0 como verdadeiro sabemos que a estatística t segue a distribuição t de Student. Portanto:

$$\begin{array}{rcccl}
P\left[t_{(\alpha/2);\nu}\right. & \leq & t & \leq & \left. t_{(1-\alpha/2);\nu}\,\right] = 1-\alpha \\
P\left[-t_{(1-\alpha/2);\nu}\right. & \leq & t & \leq & \left. t_{(1-\alpha/2);\nu}\,\right] = 1-\alpha \\
P\left[-t_{(1-\alpha/2);\nu}\right. & \leq & \frac{\overline{x}-\mu}{s/\sqrt{n}} & \leq & \left. t_{(1-\alpha/2);\nu}\,\right] = 1-\alpha \\
P\left[\left(\frac{s}{\sqrt{n}}\right)(-t_{(1-\alpha/2);\nu})\right. & \leq & \overline{x}-\mu & \leq & \left.\left(\frac{s}{\sqrt{n}}\right) t_{(1-\alpha/2);\nu}\,\right] = 1-\alpha \\
P\left[-\left(\frac{s}{\sqrt{n}}\right) t_{(1-\alpha/2);\nu} - \overline{x}\right. & \leq & -\mu & \leq & \left.\left(\frac{s}{\sqrt{n}}\right) t_{(1-\alpha/2);\nu} - \overline{x}\,\right] = 1-\alpha \\
P\left[\overline{x} - \left(\frac{s}{\sqrt{n}}\right) t_{(1-\alpha/2);\nu}\right. & \leq & \mu & \leq & \left.\overline{x} + \left(\frac{s}{\sqrt{n}}\right) t_{(1-\alpha/2);\nu}\,\right] = 1-\alpha
\end{array}$$

INTERVALO DE CONFIANÇA

$$\overline{x} \pm \left(\frac{s}{\sqrt{n}}\right) t_{(1-\alpha/2);\nu}$$

é chamado de Intervalo de Confiança de $100(1-\alpha)\%$,
onde α é o nível de significância.

Exemplo 15.B Produção em Florestas de Plantadas

Com base nas informações do exemplo anterior construa Intervalos de Confiança 95% para da produção floresta.

$$\begin{array}{rcl}
\text{Intervalo de Confiança de 95\%} & \Rightarrow & \alpha = 0.05 \Rightarrow t_{(1-\alpha/2);\nu} = t_{0.975;19} = 2.09 \\
 & \Rightarrow & \overline{x} \pm \left(\frac{s}{\sqrt{n}}\right) t_{(1-\alpha/2);\nu} \Rightarrow 280 \pm \left(\frac{25}{\sqrt{20}}\right) 2.09 \\
 & \Rightarrow & 280 \pm 11.7 \text{ st/ha} \\
 & \Rightarrow & [268.3\ ,\ 291.7] \text{ st/ha}
\end{array}$$

Interpretação:

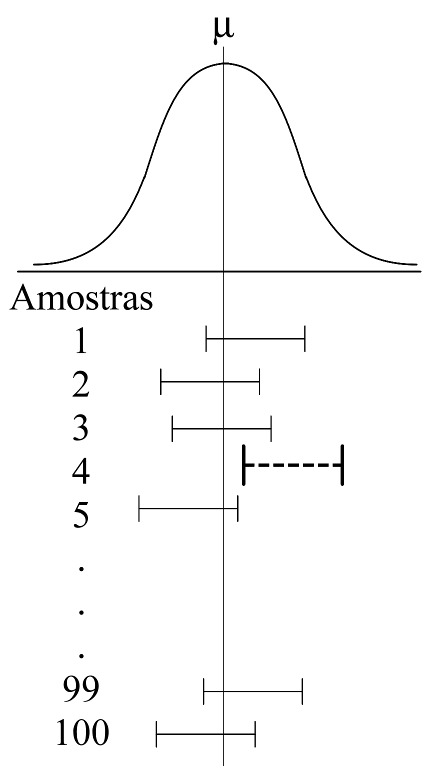
Questão 1: O I.C. de 95% contém a média populacional (μ)?

Não sabemos.

Questão 2: O I.C. de 95% significa que existe 95% de probabilidade da média populacional (μ) estar dentro do intervalo [268.3 , 291.7]?

Não! O I.C. é fixo e a média μ é fixa, logo, ou μ está dentro do intervalo ou não está!

O que é Intervalo de Confiança de 95%?



Se repertirmos o mesmo procedimento de amostragem e construção do Intervalo de Confiança várias vêzes, Esperamos que em 95% das vêzes o Intervalo de Confiança construido conterá a média populacional (μ).

Exemplo 15.C Tamanho Médio dos Peixes

Analisando os dados de uma amostra de 200 peixes pescados num reservatório, uma Eng. Florestal compilou os seguintes dados:

Variáveis	Média Amostral	Desvio Padrão Amostral
Comprimento (cm)	35	8.3
Peso (kg)	0.870	0.120

Com base em tais dados, ela deseja estimar o tamanho médio dos peixes do reservatório em relação as duas variáveis.

a. Encontre os Intervalos de Confiança de 95% para ambas variáveis.

$t_{0.975;\nu=99} =$

Variável	Média Amostral	Amplitude do Intervalo	Limite Inferior	Limite Superior
X	$\overline{x}$	$(s/\sqrt{n})t_{0.975;99}$	$\overline{x} - (s/\sqrt{n})t_{0.975;99}$	$\overline{x} + (s/\sqrt{n})t_{0.975;99}$
Comprimento (cm)				
Peso (kg)				

b. Qual variável tem o maior intervalo de confiança?

c. Como você interpreta o resultado da questão anterior?

Conceitos-Chave

TESTE DE HIPÓTESES - ERRO TIPO I - NÍVEL DE SIGNIFICÂNCIA - VALOR-P - ABORDAGEM CLÁSSICA - VALOR CRÍTICO - REGIÃO DE REJEIÇÃO - TESTE MONOCAUDAL - TESTE BICAUDAL - ESTIMATIVA PONTUAL - INTERVALO DE CONFIANÇA PARA A MÉDIA DE UMA POPULAÇÃO - INTERPRETAÇÃO DO INTERVALO DE CONFIANÇA

Exercícios

15.1 Em florestas nativas a "dominância" das espécies arbóreas é medida pela somatória das áreas seccionais das árvores (área basal). Num levantamento em fragmento florestal, as medições em 7 parcelas resultou nos seguintes dados:

Espécie	Área Basal (m^2/ha)	
	Média	Desvio Padrão
Jatobá	2.5	0.4
Embaúba	1.2	0.3
Jequitibá	5.7	1.8

a. Encontre o Intervalo de Confiança de 90% para área basal de cada uma das espécies.

b. Qual das espécies tem maior intervalo de confiança? O que significa isso?

15.2 Num estudo sobre o peso médio de capivaras, a pesquisadora encontrou o seguinte intervalo de Confiança 95%: $[4.5, 7.8]$. A pesquisadora deseja saber agora se o peso médio da população estudada é estatísticamente diferente de 5kg.

a. Estabeleça as hipótese nula e alternativa para a pergunta da pesquisadora.

b. É possível responder esta pergunta? Explique.

c. Se for possível responder, qual seria o resultado do teste de hipótese.

15.3 Uma fábrica de panéis de madeira deseja duplicar a sua produção. A duplicação acarretaria num consumo adicional de 1300000 m^3/ano de madeira. Com base no sistema de inventário da empresa, que dispõe de 1500 parcelas, o gerente florestal encontrou que as florestas da empresa têm uma capacidade excedente de produção média de 1320000 m^3/ano com desvio padrão de 170000 m^3/ano. A empresa deve fazer a duplicação? Explique.

15.4 Sabe-se que numa dada floresta tropical não perturbada a abundância de espécies "pioneiras" fica em torno de 10%. Com aumento de perturbações antrópicas, a abundância destas espécies tende a crescer. Na demarcação de uma reserva florestal de área total de 50000 ha, ecologistas e engenheiros florestais discutem a incorporação de uma área extra de 7500 ha onde o levantamento de campo (50 parcelas) revelou abundância média de espécies pioneiras de 15% (desvio padrão de 9%). O objetivo da demarcação da reserva é proteger "áreas representativas de ecossistemas locais que não sofreram influência antrópica marcante". Á área extra de 7500 ha deve ou não deve ser incorporada à reserva? Justifique estatisticamente a sua resposta.

15.5 Uma pesquisadora deseja estimar o tamanho da população de *Furnaris rufus* (joão-de-barro) numa dada localidade do interior de Minas Gerais. Num primeiro levantamento, a pesquisadora capturou 30 pássaros (n_1) e marcou-os todos com anilhas. Num segundo levantamento, a pesquisadora capturou 40 pássaros (n_2) dos quais 12 (m) possuíam a anilha referente ao primeiro levantamento. A teoria de marca-e-recaptura estabele que o tamanho da população (valor amostral) e sua variância devem ser obtidos pelos estimadores:

$$\text{Tamanho de população} \quad : \quad \widehat{N} = \frac{(n_1+1)(n_2+1)}{(m+1)} - 1$$

$$\text{Variância} \quad : \quad s^2_{\widehat{N}} = \frac{(n_1+1)(n_2+1)(n_1-m)(n_2-m)}{(m+1)^2(m+2)}$$

$$\text{Intervalo de Confiança de 95\%} \quad : \quad \widehat{N} \pm 2\sqrt{s^2_{\widehat{N}}}$$

a. Qual o intervalo de confiança de 95% para o tamanho desta população ?

b. A pesquisadora pode considerar a população de *Furnarius rufus* como tendo 100 pássaros ? Explique.

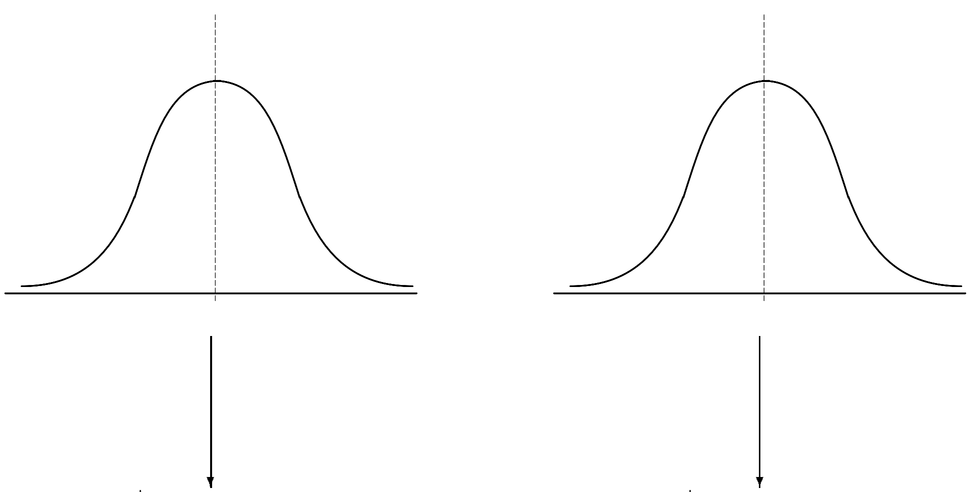
16. Comparando duas Populações: "Teste t"

Modelo Estatístio para Comparação de Duas Populações

Para compararmos duas populações devemos assumir que cada população possui distribuição Normal com suas médias e variâncias respectivas. De cada população se toma uma amostra aleatória simples, com tamanhos n_1 e n_2, não necessariamente iguais.

Poplução 1
Distribuição Normal
$(\mu_1,\ \sigma_1^2)$

Poplução 2
Distribuição Normal
$(\mu_2,\ \sigma_2^2)$



Amostra
$(\overline{x}_1,\ s_1^2,\ n_1)$

Amostra
$(\overline{x}_2,\ s_2^2,\ n_2)$

Nossa intensão é testar dois grupos de hipóteses:

Comparação de Médias:

$$H_0 : \mu_1 = \mu_2$$
$$H_\alpha : \mu_1 \neq \mu_2$$

Comparação de Variâncias:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_\alpha : \sigma_1^2 \neq \sigma_2^2$$

Comparando as Variâncias de Duas Populações

A comparação das variâncias de duas populações se baseia nas hipóteses:

Teste Bicaudal

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_\alpha : \sigma_1^2 \neq \sigma_2^2$$

Teste Monocaudal

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$\begin{cases} H_\alpha : \sigma_1^2 > \sigma_2^2 \\ \quad \text{ou} \\ H_\alpha : \sigma_1^2 < \sigma_2^2 \end{cases}$$

Para testar estas hipóteses utilizamos a estatística:

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2}$$

Sendo a hipótese nula verdadeira, temos

$$\sigma_1^2 = \sigma_2^2 = \sigma^2$$

e a estatística F fica:

$$F = \frac{s_1^2/\sigma_1^2}{s_2^2/\sigma_2^2} = \frac{s_1^2/\sigma^2}{s_2^2/\sigma^2} = \frac{s_1^2}{s_2^2}$$

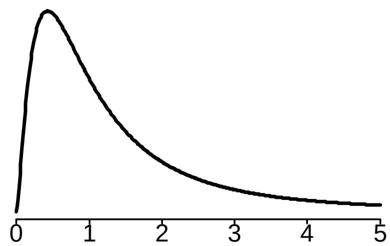
Se H_0 é verdadeira, ambas amostras foram obtidas de duas populações "Normais" com igual variância (σ^2).

Nesta situação a estatística F tem Distribuição F, a qual tem as seguintes propriedades:

- F só assume valores positivos, pois é uma razão de duas variâncias ($0 < F < +\infty$);
- Possui dois parâmetros:

  $\nu_1 = n_1 - 1 \Rightarrow$ Graus de Liberdade do Numerador

  $\nu_2 = n_2 - 1 \Rightarrow$ Graus de Liberdade do Denominador
- A distribuição F é unimodal assimétrica a direita, tendo o gráfico:



- O Anexo D apresenta uma tabela da distribuição F para alguns valores de graus de liberdade para o nível de significância de 5%.

Exemplo 16.A Estimando a Distância de Animais Observados

Um pesquisador deseja testar dois métodos para determinar a distância do observador ao animal em levantamento de fauna. Durante um levantamento, a distância para cada animal observado foi medida utilizando os dois métodos obtendo-se os seguintes resultados:

Método	Distâncias					Variância
Utilizando Rangefinder	25	30	16	13	21	46.5
Utilizando Trena	20	31	18	10	23	58.3

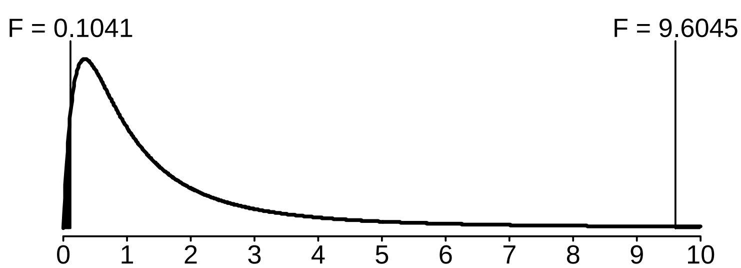
- Hipóteses testadas:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_\alpha : \sigma_1^2 \neq \sigma_2^2$$

- Estatística F amostral:

$$F = \frac{46.5}{58.3} = 0.798 \quad \textbf{ou} \quad F = \frac{58.3}{46.5} = 1.254$$

- Para o nível de significância de 5% ($\alpha = 0.05$) com graus de liberdade $\nu_1 = 4$ e $\nu_2 = 4$ temos:



- Região de Rejeição:
    * O quantil 2.5% é 0.104 e o quantil 97.5% é 9.605;

* Os valores da estatística F observada são menos extremos que os quantis;

* Não se rejeita H_0.

Teste F Bicaudal

Note que os valores na tabela do Anexo D apresentam apenas um valor (quantil 95% ou 97.5%). Para simplificar os cálculos e o teste de hipótese, devemos utilizar o seguinte procedimento:

1. Calcular a estatística F sempre na forma:

$$\boxed{F = \frac{\max\{s_1^2; s_2^2\}}{\min\{s_1^2; s_2^2\}}}$$

2. Rejeitar H_0 quando o valor da estatística for maior que o quantil $(1-\alpha/2)100\%$:

$$\boxed{F \geq F_{1-\alpha/2;\nu_1;\nu_2}}$$

Exemplo 16.B Comparando Processos de Produção de Papel

Uma fábrica de papel deseja introduzir um novo processo de fabricação que é considerado mais eficiente tanto em termos de custo com em termos ambientais. A Engenheira Florestal responsável decidiu fazer um teste comparando o novo processo contra o processo tradicional em termos de *gramatura*, que é o peso do metro quadrdado de papel, obtendo os seguintes resultados:

PROCESSO	GRAMATURA (g/m^2)						n	s^2
Tradicional	120	140	80	75	110	150	6	937.5
Novo	105	95	108	120	90		5	137.3

- Hipóteses:

$$H_0 : \sigma_{\mathrm{T}}^2 = \sigma_{\mathrm{N}}^2$$
$$H_\alpha : \sigma_{\mathrm{T}}^2 > \sigma_{\mathrm{N}}^2$$

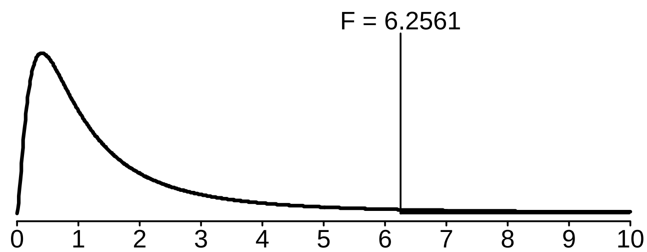
- Estatística F amostral:

$$F = \frac{937.5}{137.3} = 6.828$$

- Região de Rejeição:

  Sob H_0, a estatística F terá distribuição F com graus de liberdade $\nu_1 = 5$ e $\nu_2 = 4$. O valor do quantil 95% neste caso é 6.256 (veja Anexo D).

  Gráfico:



- Como $6.8281 > 6.256$ ($F > F_{0.95;5;4}$), rejeita-se H_0 e conclui-se que o método tradicional tem variância maior que o novo.

<u>*Teste F Monocaudal*</u>

No caso do teste F monocaudal, devemos utilizar o seguinte procedimento:

1. Calcular a estatística F segundo a forma como a hipótese alternativa é estabelecida:

$$\boxed{\begin{array}{l} H_\alpha : \sigma_1^2 > \sigma_2^2 \Longrightarrow F = \dfrac{s_1^2}{s_2^2} \\ H_\alpha : \sigma_1^2 < \sigma_2^2 \Longrightarrow F = \dfrac{s_2^2}{s_1^2} \end{array}}$$

2. Rejeitar H_0 quando o valor da estatística for maior que o quantil $(1-\alpha)100\%$:

$$\boxed{F \geq F_{1-\alpha;\nu_1;\nu_2} \quad .}$$

Comparação de Médias

A comparação de médias geralmente envolve as hipóteses:

$$H_0 : \mu_1 = \mu_2 \Rightarrow \mu_1 - \mu_2 = 0$$
$$H_\alpha : \mu_1 \neq \mu_2 \Rightarrow \mu_1 - \mu_2 \neq 0$$

O que implica que as médias dos tratamentos são iguais.

Uma forma mais genérica de se estabelecer estas hipóteses é afirmar que as médias diferem uma certa quantia θ, onde $\theta = 0$ é uma das possibilidades:

$$\begin{aligned} H_0 &: \mu_1 - \mu_2 = \theta \\ H_\alpha &: \mu_1 - \mu_2 \neq \theta \end{aligned}$$

A estatística utilizada para testar estas hipóteses é

$$\boxed{t = \frac{(\overline{x}_1 - \overline{x}_2) - \theta}{\sqrt{\widehat{\text{Var}}(\overline{x}_1 - \overline{x}_2)}}}$$

A estatística t assim calculada dependerá de como a estimativa da *variância da diferença das média* ($\widehat{\text{Var}}(\overline{x}_1 - \overline{x}_2)$) é obtida. Como as amostras que geraram $\overline{x}_1$ e $\overline{x}_2$ são independentes temos que:

$$\text{Var}(\overline{x}_1 - \overline{x}_2) = \text{Var}(\overline{x}_1) + \text{Var}(\overline{x}_2)$$

- Se $\sigma_1^2 = \sigma_2^2$, então

  $$\text{Var}(\overline{x}_1 - \overline{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \left(\frac{1}{n_1} + \frac{1}{n_2}\right)\sigma^2$$

  Neste caso a estimativa de σ^2 deve ser uma composição de σ_1^2 e σ_2^2:

  $$\boxed{\begin{aligned} \widehat{\text{Var}}(\overline{x}_1 - \overline{x}_2) &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_{\text{C}}^2 \\ s_{\text{C}}^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \end{aligned}}$$

- Se $\sigma_1^2 \neq \sigma_2^2$, então

  $$\text{Var}(\overline{x}_1 - \overline{x}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

  e a estimativa fica:

  $$\boxed{\widehat{\text{Var}}(\overline{x}_1 - \overline{x}_2) = \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

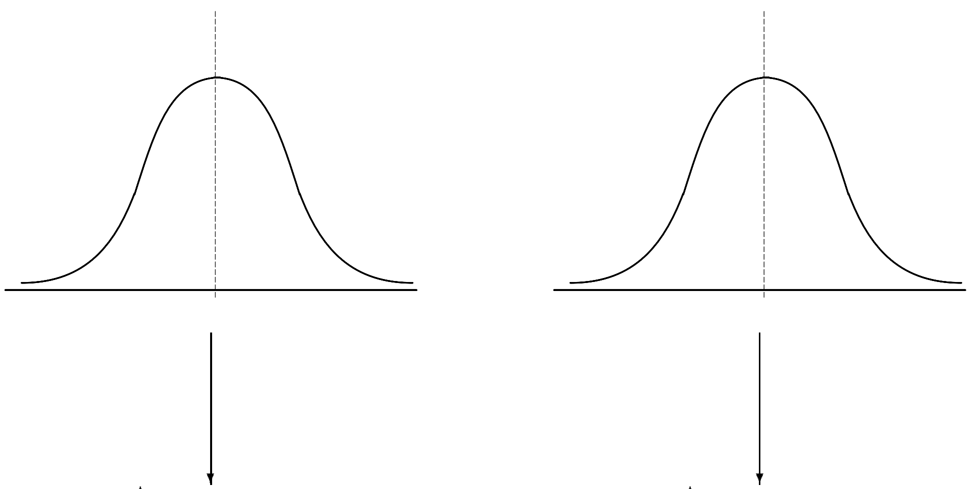
Contexto Observacional versus Contexto Experimental

A comparação de médias têm signficado diferente se
os dados foram obtidos num contexto de
Pesquisa Observacional ou
Pesquisa Experimental.

Os exemplos considerados até agora se referem a Pesquisa Observacional, pois tinhamos **duas** populações distintas das quais se tomou amostras independentes.

Poplução 1
Distribuição Normal
$(\mu_1,\ \sigma_1^2)$

Poplução 2
Distribuição Normal
$(\mu_2,\ \sigma_2^2)$



Amostra
$(\overline{x}_1,\ s_1^2,\ n_1)$

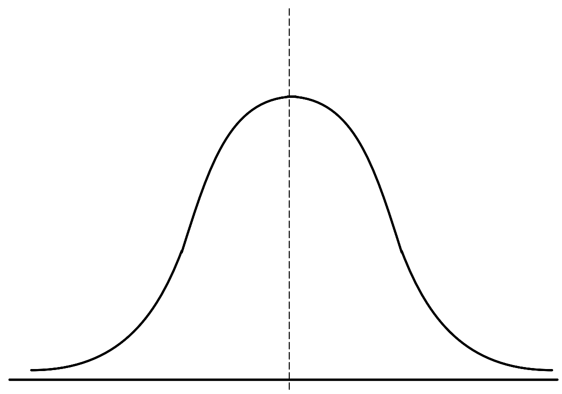
Amostra
$(\overline{x}_2,\ s_2^2,\ n_2)$

O contexto da Pesquisa Experimental surge quando o pesquisador parte de uma **única** população , toma **uma** amostra e aplica nesta amostra **dois** diferentes tipos de manipulação experimental.

O contexto experimental, portanto, implica em:

1. As Unidades Experimentais $(n_1 + n_2)$ foram tomadas aleatoriamente de uma **mesma** população.
2. Os Tratamentos (manipulação experimental das unidades) foram atribuidos de *modo aleatório* a cada uma das unidades experimentais.
3. Os tratamentos foram repetidos em mais de uma undade experimental.
4. A aplicação do tratamento em uma das unidades experimentais é completamente independente da aplicação dos tratamentos nas demais unidades.

Poplução Original
Distribuição Normal
$(\mu,\ \sigma^2)$



Amostra Aleatória
$(n_1 + n_2)$

Tratamento 1
$(\overline{x}_{1},\ s_{1}^{2},\ n_1)$

Tratamento 2
$(\overline{x}_{2},\ s_{2}^{2},\ n_2)$

Note que no contexto experimental, a rejeição da hipótese $H_0 : \sigma_1^2 - \sigma_2^2$ já implica que os tratamentos aplicados são diferentes.

Consideraremos agora como o teste de médias deve ser aplicado em diferentes situações de pesquisas observacionais e experimentais.

Teste t para Duas Médias: Observações Pareadas

As observações são ditas pareadas quando feitas
num mesmo objeto ou agrupada em pares no campo.

Desta forma a **diferença** entre duas observações pareadas tem significado prático e podemos pensar na **média** e **variância** das diferenças entre observações.

A hipótese nula, neste caso, pode ser transformada para:

$$H_0 : \mu_1 = \mu_2 \Longrightarrow H_0 : \mu_1 - \mu_2 = 0 \Longrightarrow H_0 : \mu_{\mathrm{d}} = 0$$

Novamente, podemos tornar esta hipótese mais genérica:

$$H_0 : \mu_{\mathrm{d}} = \theta$$
$$H_\alpha : \mu_{\mathrm{d}} \neq \theta$$

Assim, o teste t para observações pareadas se resume ao teste t para uma população testando hipótese nula da **média diferenças** ser igual a uma certa quantia θ.

$$t = \frac{\overline{d} - \theta}{\sqrt{s_{\mathrm{d}}^2/n}}$$

Exemplo 16.C Estimando a Distância de Animais Observados - II

Voltando ao exemplo da medição da distância observdor-animal, podemos re-organizar a tabela na forma:

MÉTODO	DISTÂNCIAS					$\overline{x}$	s^2
Utilizando Rangefinder	25	30	16	13	21	21.0	46.5
Utilizando Trena	20	31	18	10	23	20.4	58.3
Diferença	5	-1	-2	3	-2	0.6	10.3

- Hipóteses:

$$H_0 : \mu_{\mathrm{R}} - \mu_{\mathrm{T}} = 0$$
$$H_\alpha : \mu_{\mathrm{R}} - \mu_{\mathrm{T}} \neq 0$$

- Estatística:

$$t = \frac{0.6}{\sqrt{10.3/5}} = 0.418$$

- Região de Rejeição:

$$|t| \geq t_{0.975;4}$$
$$|t| \geq 2.776$$

- Conclusão: Como $|0.418| < 2.776$, não se rejeita H_0. Conclui-se que não há evidências de que os dois métodos sejam diferentes entre si.

Teste t para Duas Médias: Variâncias Homogêneas

Antes de compararmos duas médias onde as observações **não** foram pareadas, devemos primeiro comparar as variâncias através do teste F, testando as hipóteses:

$$H_0 : \sigma_1^2 = \sigma_2^2$$
$$H_\alpha : \sigma_1^2 \neq \sigma_2^2$$

Se H_0 não for rejeitada, assumimos que as variâncias são iguais e comparamos as médias através da estatística:

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - \theta}{\sqrt{\widehat{\text{Var}}(\overline{x}_1 - \overline{x}_2)}}$$

onde a variância da diferença das médias é obtida por:

$$\widehat{\text{Var}}(\overline{x}_1 - \overline{x}_2) = \left(\frac{1}{n_1} + \frac{1}{n_2}\right) s_{\text{C}}^2$$
$$s_{\text{C}}^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Sob H_0, a estatística t calculada seguirá a distribuição t de Student com graus de liberdade $\nu = n_1 + n_2 - 2$.

Exemplo 16.D Procedências de Pinus oocarpa

Um Engenheiro Florestal deseja saber se uma procedência mais produtiva de *Pinus oocarpa* (procedência A) difere da procedência menos produtiva (procedência B) em pelo menos 10 st/ha.ano. Os dados obtidos foram:

Procedência	Incremento Médio ($st/ha.nao$)								$\overline{x}$	s^2
A	45.6	42.1	44.9	45.1	47.6	46.7	45.5	48.9	45.8	4.1114
B	30.1	21.6	27.6	27.3	30.4	31.4	34.1	30.6	29.1	13.8512

Testando igualdade de variâncias

- Hipóteses:

$$H_0 : \sigma^2_{\text{A}} = \sigma^2_{\text{B}}$$
$$H_\alpha : \sigma^2_{\text{A}} \neq \sigma^2_{\text{B}}$$

- Estatística:

$$F = \frac{13.8512}{4.1114} = 3.369$$

- Região de Rejeição:

$$F \geq F_{0.975;7;7} \Rightarrow F \geq 4.995$$

- Conclusão: Como $3.369 < 4.995$ não se rejeita H_0. Assume-se que as variâncias são iguais.

Comparando as Médias

- Hipóteses:

$$H_0 : \overline{x}_{\text{A}} - \overline{x}_{\text{B}} = 10$$
$$H_\alpha : \overline{x}_{\text{A}} - \overline{x}_{\text{B}} > 10$$

- Estatística:

$$\begin{aligned}
\sigma^2_1 = \sigma^2_2 \Rightarrow \widehat{\text{Var}}(\overline{x}_1 - \overline{x}2) &= \left(\frac{1}{n_1} + \frac{1}{n_2}\right) \frac{(n_1 - 1)s^2_1 + (n_2 - 1)s^2_2}{n_1 + n_2 - 2} \\
\widehat{\text{Var}}(\overline{x}_1 - \overline{x}2) &= \left(\frac{1}{8} + \frac{1}{8}\right) \frac{(8-1)4.11114 + (8-1)13.8512}{8+8-2} = 2.2453 \\
t &= \frac{(\overline{x}_{\text{A}} - \overline{x}_{\text{B}}) - \theta}{\sqrt{\widehat{\text{Var}}(\overline{x}_1 - \overline{x}2)}} = \frac{(45.8 - 29.1) - 10}{\sqrt{2.2453}} = \frac{6.7}{1.4984} = \boxed{4.471}
\end{aligned}$$

- Região de Rejeição (teste monocaudal):

$$t \geq t_{1-\alpha;2(n-1)} \Rightarrow t \geq t_{0.95;14} \Rightarrow t \geq 1.761$$

- Conclusão: Como $4.471 > 1.761$, rejeita-se H_0 ao nível de 5% de probabilidade. Conclui-se há evidências para se afirmar que a procedência mais produtiva excede a menos produtiva em mais de 10 $st/ha.ano$.

Teste t para Duas Médias: Variâncias Heterogêneas

Quando rejeitamos H_0 no teste de comparação das variâncias,
temos que assumir que $\sigma_1^2 \neq \sigma_2^2$ e a hipótese nula
para as médias é testada utilizando a mesma estatística t.

O comportamento da estatística t sob H_0, no entanto,
dependerá se o tamanho de amostra para as duas médias
é igual ou diferente.

Amostras de Mesmo Tamanho ($n_1 = n_2$)

Para amostras de mesmo tamanho $n_1 = n_2 = n$,
a estatística t se torna:

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - \theta}{\sqrt{\frac{s_1^2 + s_2^2}{n}}}$$

Sob H_0, a estatística t terá distribuição t
de Student com graus de liberdade $\nu = 2(n-1)$.

Amostras de Tamanhos Diferentes ($n_1 \neq n_2$)

No caso das amostras terem tamanhos diferentes, a estatística t fica:

$$t = \frac{(\overline{x}_1 - \overline{x}_2) - \theta}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

a qual terá distribuição t de Student sob H_0
com graus de liberdade igual a:

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$

Exemplo 16.E Métodos de Resinagem

Uma Engenheira Florestal testou dois métodos de resinagem em matrizes de *Pinus elliottii*. Um grupo de 18 das melhores matrizes foi selecionado. Através de sorteio aleatório, aplicou-se em 9 matrizes o tratamento com ácido sufúrico a 30% (tratamento A), enquanto que as demais 9 matrizes receberam o tratamento de ácido sufúrico a 15% (Tratamento B). Os resultados obtidos foram:

TRAT.	PRODUÇÃO DE RESINA(g)									$\overline{x}$	s^2
A	2326	2206	1835	1434	1629	1761	1511	2146	1548	1821.778	108740.944
B	6006	3455	3115	3376	2609	2582	3674	2648	2012	3275.222	1324733.194

Testando a igualdade das variâncias

- Hipóteses:

$$H_0 : \sigma^2_{\text{A}} = \sigma^2_{\text{B}}$$
$$H_\alpha : \sigma^2_{\text{A}} \neq \sigma^2_{\text{B}}$$

- Estatística:

$$F = \frac{1324733.194}{108740.944} = 12.182$$

- Região de Rejeição:

$$F \geq F_{0.975;8;8} \Rightarrow F \geq 4.357$$

- Conclusão: Como $12.182 > 4.357$, rejeita-se H_0. Assume-se que as variâncias são diferentes.

Comparando as Médias

- Hipóteses:

$$H_0 : \overline{x}_{\text{A}} = \overline{x}_{\text{B}}$$
$$H_\alpha : \overline{x}_{\text{A}} \neq \overline{x}_{\text{B}}$$

- Estatística:

$$\begin{aligned}
\widehat{\text{Var}}(\overline{x}_1 - \overline{x}2) &= \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = \frac{1324733.194}{9} + \frac{108740.944}{9} = 159274.9034 \\
t &= \frac{(\overline{x}_{\text{A}} - \overline{x}_{\text{B}}) - \theta}{\sqrt{\widehat{\text{Var}}(\overline{x}_1 - \overline{x}2)}} = \frac{(1821.778 - 3275.222) - 0}{\sqrt{159274.9034}} = \frac{-1453.444}{399.0926} = \boxed{\text{-3.642}}
\end{aligned}$$

- Região de Rejeição (teste bicaudal):

$$|t| \geq t_{1-\alpha/2;2(n-1)} \Rightarrow |t| \geq t_{0.975;16} \Rightarrow |t| \geq 2.120$$

- Conclusão: Como $|-3.642| > 2.120$, rejeita-se H_0 ao nível de 5% de probabilidade. Conclui-se há evidências para se afirmar que os métodos de resinagem são diferentes.

Exemplo 16.F Comparando Processos de Produção de Papel - II

Voltando ao exemplo da fábrica de papel, onde foi rejeitada a hipótese de igualdade das variâncias, a Engenheira Florestal deseja agora testar a hipótese sobre as médias:

$$H_0 : \mu_{\mathrm{T}} - \mu_{\mathrm{N}} = 0$$
$$H_\alpha : \mu_{\mathrm{T}} - \mu_{\mathrm{N}} \neq 0$$

Os resultados obtidos foram:

Processo	Gramatura (g/m^2)						n	$\overline{x}$	s^2
Tradicional	120	140	80	75	110	150	6	112.5	937.5
Novo	105	95	108	120	90		5	103.6	137.3

- Estatística:

$$\begin{aligned}
\widehat{\text{Var}}(\overline{x}_1 - \overline{x}2) &= \frac{937.5}{6} + \frac{137.3}{5} = 183.711 \\
t &= \frac{112.5 - 103.6}{\sqrt{183.711}} = \frac{8.90}{13.554} = \boxed{0.657}
\end{aligned}$$

- Região de Rejeição:

$$\begin{aligned}
\nu &= \left(\frac{937.5}{6} + \frac{137.3}{5}\right)^2 / \left[\frac{\left(\frac{937.5}{6}\right)^2}{6-1} + \frac{\left(\frac{137.3}{5}\right)^2}{5-1}\right] \\
&= \frac{33749.3641}{5071.3254} = 6.655 \rightarrow 6 \\
|t| &\geq t_{0.975;6} \Rightarrow |t| > 2.447
\end{aligned}$$

- Conclusão: Como $|0.657| < 2.447$, não se rejeita H_0 e conclu-se que não há evidência (ao nível de 5% de probabilidade) para afirmar que os processos produzem papel com gramaturas médias diferentes.

Conceitos-Chave

ESTATÍSTICA F - DISTRIBUIÇÃO F - REGIÃO DE REJEIÇÃO NA DISTRIBUIÇÃO F - TESTE F BICAUDAL - TESTE F MONOCAUDAL - ESTATÍSTICA t PARA COMPARAÇÃO DE DUAS MÉDIAS - VARIÂNCIA DA DIFERENÇA DAS MÉDIAS - CONTEXTO OBSERVACIONAL - CONTEXTO EXPERIMENTAL TESTE t PAREADO - TESTE t PARA VARIÂNCIAS HOMOGÊNEAS - TESTE t VARIÂNCIAS HETEROGÊNEAS COM AMOSTRAS DE MESMO TAMANHO - TESTE t VARIÂNCIAS HETEROGÊNEAS COM AMOSTRAS DE TAMANHOS DIFERENTES

Exercícios

16.1 Acredita-se que a adubação de cobertura em *Eucalyptus grandis* nem sempre produz um ganho na produção mas aumenta a homogeneidade das árvores. Formule hipóteses estatísticas apropriadas e teste-as utilizando os dados da tabela abaixo.

DAP (cm)										
Floresta Adubada										
14.9	18.6	16.8	14.6	13.8	20.7	15.8	20.1	19.4	18.4	18.5
15.3	16.6	18.9	18.1	14.4	14.5	14.0	16.3	17.6	17.6	
Floresta Não Adubada										
21.6	25.8	18.7	16.7	23.1	14.6	6.4	12.2	31.5	4.1	11.5
27.2	8.0	22.5	25.1	16.8	17.4	20.7	14.8	15.8	9.3	

16.2 Um pesquisador deseja verificar se o melhoramento genético produziu redução marcante na grau de rachadura de topo de *Eucalyptus saligna*. Comparando dois grupos de árvores o pesquisador obteve os dados na tabela abaixo. Teste as hipóteses apropriadas e estabeleça as suas conclusões.

Número de Rachaduras / cm^2								
Árvores não melhoradas								
0.2	6.6	8.0	2.2	1.9	6.8	1.7	6.3	1.4
Árvores melhoradas								
2.6	4.6	4.8	4.3	4.4	3.1	3.8		

16.3 Um engenheiro florestal deseja comparar a dominância de duas espécie nativas. Num levantamento com 31 parcelas, a espécie A teve dominância média de 5.3 m^2/ha (desvio padrão de 1.2) e a espécie B teve média de 6.7 m^2/ha (desvio padrão de 2.1). Estabeleça a hipóteses estatísticas apropriadas. Teste as hipóteses e estabeleça as suas conclusões.

16.4 Numa fábrica de celulose o engenheiro responsável notou um aumento no consumo dos produtos químicos na útlima semana. Suspeitando que este aumento se deve ao fato da madeira dos lotes mais recentes possuir densidade maior que a madeira dos lotes mais antigos, ele verificou as tabelas de controle de densidade da madeira feita no pátio da fábrica (tabela abaixo). O aumento de produtos químicos no processo de fabricação poderia ser atribuido a um aumento na densidade da madeira entrando na fábrica ?

Densidade da Maderia (g/cm^3)									
Última Semana									
0.645	0.711	0.705	0.549	0.676	0.662	0.709	0.637	0.711	0.728
0.575	0.613	0.880	0.625	0.684	0.622	0.579	0.490	0.529	0.626
0.619	0.696	0.642	0.687	0.603	0.708	0.667	0.425	0.665	
Semana Anterior									
0.856	0.704	0.755	0.582	0.836	0.550	0.819	0.741	0.681	0.713
0.882	0.844	0.601	0.627	0.628	0.691	0.599	0.669	0.746	0.607
0.547	0.680	0.654	0.701	0.741	0.563	0.723	0.719	0.675	0.620

16.5 Com o objetivo de verificar o efeito da exposição do solo sobre a microfauna do solo, uma pesquisadora fez levantamentos de microfauna antes e depois da exposição do solo em áreas desmatadas de vários tipos de ecossistemas (tabela abaixo). Teste as hipóteses apropriadas e estabeleça a sua conclusão.

Ecossistema	No. de Micro-organismos/cm^3	
	Antes	Depois
Campo Limpo	1430	780
Cerrado	2500	1020
Restinga	732	640
Caatinga	640	680
Floresta Estacional	10530	2520
Floresta Pluvial	21883	2302

17. Modelo vs. Realidade: "Teste de Qui-Quadrado"

Comparação de Frequências

Já foram vistos vários testes para comparar parâmetros de
duas populações (média e variância).

Entretanto, muitas vêzes desejamos comparar *frequências* observadas para testar se
elas são provenientes de uma mesma população

O teste de Qui-Quadrado (teste χ^2) foi construido
para comparar dois grupos de frequências.

Exemplo 17.A Distribuição dos Diâmetros das Árvores numa Floresta

Um engenheiro florestal deseja saber se a distribuição Normal é um bom modelo estatístico para os diâmetros das árvores de uma floresta plantada de *Eucalyptus camaldulensis*.

Classes de Diâmetros (cm)		Frequência
Lim. Inferior	Lim. Superior	(arv/ha)
10.0	12.5	4
12.5	15.0	23
15.0	17.5	50
17.5	20.0	132
20.0	22.5	224
22.5	25.0	286
25.0	27.5	295
27.5	30.0	243
30.0	32.5	161
32.5	35.0	73
35.0	37.5	9
		1500

A tabela acima apresenta a frequência *observada* para cada classe de diâmetro. As hipóteses sendo testadas são:

$$H_0: \quad f_i = \widehat{f_i}$$
$$H_\alpha: \quad f_i \neq \widehat{f_i} \text{ (para pelo menos um dos } i\text{)}$$

onde f_i é a frequência observada e $\widehat{f}_i$ é a frequência que esperamos para dados com distribuição Normal.

Para verificarmos se as frequências observadas e as frequências esperadas estão próximas, devemos calcuar as *frequências esperadas*. Para cada classe de diâmetro fazemos os seguintes cálculos:

Valores da Normal Padronizada para os Limites: $z_{I,i} = \dfrac{dap_{I,i} - \overline{dap}}{\sqrt{s^2_{\text{dap}}}}$

$z_{S,i} = \dfrac{dap_{S,i} - \overline{dap}}{\sqrt{s^2_{\text{dap}}}}$

Probabilidade por Classe (Tabela): $\widehat{p}_i = P(Z < z_{I,i}) - P(Z < z_{S,i})$

Frequência Esperada (N=1500): $\widehat{f}_i = \widehat{p}_i\, N$

Classes (i)	Diâmetros (cm) Limite Inferior $dap_{I,i}$	Limite Superior $dap_{S,i}$	Frequência Observada (arv/ha) f_i	Normal Padronizada Limite Inferior $z_{I,i}$	Limite Superior $z_{S,i}$	Probab. por Classe $\widehat{p}_i$	Frequência Esperada (arv/ha) $\widehat{f}_i$
1	10.0	12.5	4	-3.28	-2.74	0.0026	3.90
2	12.5	15.0	23	-2.74	-2.20	0.0108	16.20
3	15.0	17.5	50	-2.20	-1.67	0.0336	50.40
4	17.5	20.0	132	-1.67	-1.13	0.0817	122.55
5	20.0	22.5	224	-1.13	-0.59	0.1484	222.60
6	22.5	25.0	286	-0.59	-0.05	0.2025	303.75
7	25.0	27.5	295	-0.05	0.48	0.2043	306.45
8	27.5	30.0	243	0.48	1.02	0.1617	242.55
9	30.0	32.5	161	1.02	1.56	0.0945	141.75
10	32.5	35.0	73	1.56	2.10	0.0415	62.25
11	35.0	37.5	9	2.10	2.63	0.0136	20.40
			1500			0.9952	1492.8

A estatística utilizada para comparar as frequência é

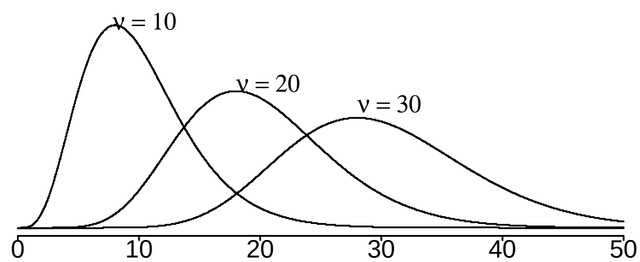
$$X^2 = \sum_{i=1}^{k} \frac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i}$$

que só poderá assumir valores positivos. Grandes valores de X^2 indicam que as frequências são bastante distintas, levando-nos a rejeitar H_0.

Sob H_0, a estatística X^2 segue a distribuição de Qui-Quadrado (χ^2). Esta distribuição tem as seguintes propriedades:

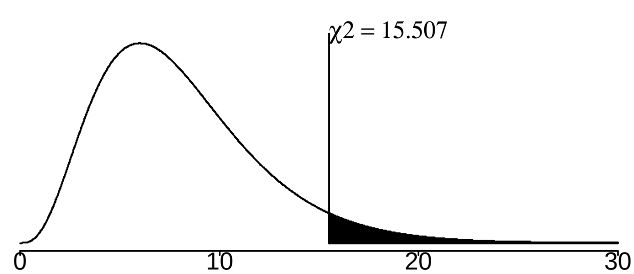
- χ^2 só assume valores positivos;

- a distribuição é unimodal, assimétrica à esquerda;
- possui um único parâmetro: *graus de liberadade* $\nu = n - 1 - p$, onde p é o número de parâmetros estimados para encontrar as frequências esperadas.



Classes (i)	Diâmetros (cm) Limite Inferior $dap_{I,i}$	Limite Superior $dap_{S,i}$	Frequência Observada (arv/ha) f_i	Frequência Esperada (arv/ha) $\widehat{f}_i$	X^2 $\frac{(f_i - \widehat{f}_i)^2}{\widehat{f}_i}$
1	10.0	12.5	4	3.90	0.003
2	12.5	15.0	23	16.20	2.854
3	15.0	17.5	50	50.40	0.003
4	17.5	20.0	132	122.55	0.729
5	20.0	22.5	224	222.60	0.009
6	22.5	25.0	286	303.75	1.037
7	25.0	27.5	295	306.45	0.428
8	27.5	30.0	243	242.55	0.001
9	30.0	32.5	161	141.75	2.614
10	32.5	35.0	73	62.25	1.856
11	35.0	37.5	9	20.40	6.371
			1500	1492.8	15.905

A valor crítico para χ^2 ao nível de 5% de probabilidade e com $\nu = 11 - 1 - 2 = 8$ graus de liberdade é 15.507. Como



$$X^2 = 15.905 > \chi^2_{0.95;\nu=8} = 15.507$$

rejeitamos H_0 e concluímos que há evidências (ao nível de 5% de probabilidade) de que o diâmetro das árvores da floresta não segue a distribuição Normal.

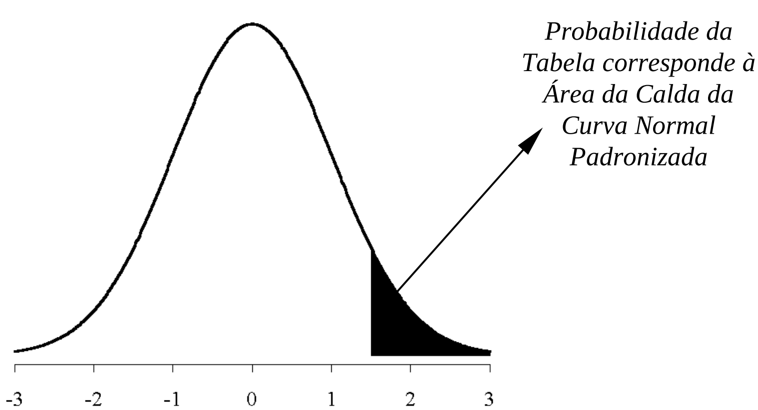
Anexo A: Tabela de Números Aleatórios

Linhas	Colunas									
	1	2	3	4	5	6	7	8	9	10
1	37330	53935	13353	39148	90771	13159	33960	06997	84693	46795
2	10556	08243	70946	57339	72234	59692	52999	93494	95782	09651
3	45811	96068	33049	45243	27243	94492	67854	01933	75224	77355
4	73736	58198	34805	92339	46743	87758	53546	91205	20339	80666
5	17223	12221	61507	77114	72285	69923	05809	06493	97064	46339
6	81548	95763	83937	91219	01721	17329	93669	60293	03667	21710
7	90216	94892	06615	01442	53108	21428	83671	71161	47712	73850
8	00764	91156	09068	03279	05814	60628	50454	75238	85013	03935
9	20759	16044	24367	90989	46036	74737	77971	76094	96855	71792
10	42933	66329	75711	94739	36532	97440	27111	15222	30748	68995
11	90227	25813	47260	18871	32378	83526	45759	91977	17721	94625
12	43376	01076	85915	30545	47255	04452	24526	77645	13186	42302
13	63024	89093	09505	22752	03459	61945	92030	77252	68334	14381
14	57417	62107	69764	44870	58163	36237	50600	19644	17087	57247
15	42252	61257	07265	89677	80135	78227	41473	70922	67684	39150
16	78009	80035	01594	31923	16715	13916	58606	83282	25382	04772
17	38594	79867	81298	27956	23186	58969	40255	53449	06430	02044
18	61166	10972	89696	72737	74575	33628	07724	09325	55122	03607
19	96847	09658	78425	27341	01678	72410	07312	18256	14849	41021
20	96233	37435	02304	86004	36919	34869	36691	05771	70516	23301
21	96723	28815	56751	06769	79709	94181	64349	87804	51772	20745
22	83778	80486	07191	27820	12450	50519	12220	01604	42242	30684
23	81870	45840	40560	06232	48136	70977	79352	60376	65229	78414
24	79836	38814	35247	52688	54767	59184	09034	45814	08280	11287
25	87161	71681	12125	66883	61697	17181	54093	56464	11199	07275
26	87783	75633	05979	57992	39071	52546	62850	73274	29428	35256
27	08119	21105	09347	34139	56304	12470	98466	22863	62609	36216
28	18922	09414	30267	31239	05542	90793	48181	04769	37248	31929
29	17629	55417	41930	85312	67452	19141	27944	78229	88922	66229
30	85876	53986	94018	97123	26576	85913	14366	63907	19704	24935
31	39992	84306	70490	49481	78479	81102	99329	90414	53003	74881
32	41986	50658	70507	53404	25900	28854	61275	59667	51324	67022
33	99382	46409	04296	92514	19794	81119	23260	64028	61763	50013
34	72215	07414	94276	26246	50979	22571	64260	32572	97833	10481
35	35633	43596	50301	44167	92491	03588	41259	10595	59500	78197
36	88312	90675	83843	39813	81547	56037	85139	86009	98724	00114
37	97936	08967	06297	90721	11649	00469	90058	19169	42794	65151
38	70453	97653	82137	31617	82240	78879	99752	33814	27956	00884
39	13199	73340	29241	96338	78479	01993	83968	45405	24145	99657
40	86790	40381	32919	51971	02988	45995	01324	43030	76348	11470
41	64480	68876	88985	23621	77087	38588	90692	19155	47371	14490
42	68173	57677	77372	71297	77786	48919	69894	25420	45816	11371
43	03398	42731	22387	81826	28712	38888	20897	85387	09919	27589
44	92656	62471	96454	40858	51211	86127	24676	84158	09159	52302
45	03659	61742	79474	10843	48114	15666	38553	05251	47571	80037
Linhas	Colunas									
	11	12	13	14	15	16	17	18	19	20

1	42165	53648	23781	61440	07050	37998	91935	77759	19704	74933
2	17186	35082	11609	21142	49037	59028	60429	64920	16096	86919
3	88019	44552	01298	92145	22778	16430	75500	03045	05800	10042
4	73984	06982	57401	46327	33441	39177	73562	34604	94405	54252
5	64178	29530	11201	11848	23831	35181	71664	13627	47908	38205
6	60987	63701	64052	03482	15119	67928	85876	28907	78476	78618
7	62759	39064	69267	45633	21715	51964	24053	42376	17255	74126
8	69098	47903	57767	00789	00197	94958	91347	17714	12752	71803
9	09690	10998	81103	29359	73418	53278	41076	48164	20004	07060
10	34934	04670	94135	30083	42034	02213	21936	03303	92272	00319
11	76303	96793	01008	57706	93717	15862	00642	81828	73887	76319
12	02455	23467	51836	46352	71524	91560	69802	22287	57267	87487
13	83810	82163	28566	15985	28592	67643	96112	45673	35587	89684
14	27329	89043	60936	61771	73344	97478	49700	09037	64897	34947
15	41187	42664	02228	81133	27379	08643	74930	54711	03608	91965
16	74140	30401	40969	24053	03853	61559	05995	24126	17948	14561
17	69008	40976	23097	67454	08611	12703	25426	51482	10591	43370
18	70862	79284	88385	87849	67488	23064	17334	92617	97037	01607
19	93192	64786	14476	06706	88219	33105	13158	03142	67513	39149
20	16146	24939	79222	31585	52940	01301	70503	82704	39142	12421
21	16218	64385	76563	55292	44910	31658	27920	14819	12959	90258
22	50985	84421	69289	09914	85585	88017	02160	03604	84955	53730
23	88210	60901	00889	36179	95595	97065	91100	93526	73198	12242
24	26357	51990	46249	62216	85683	21568	33080	05738	38108	66755
25	16629	91364	73398	13823	19038	40454	96427	90364	18166	55014
26	89864	82058	23929	00999	69219	39167	12819	63579	68274	95242
27	41764	11638	98271	81056	78376	32132	91701	82895	30675	24803
28	49543	58827	21371	75317	91602	48134	09999	08958	70837	57100
29	92900	95237	86410	15652	21033	43878	43415	77988	22126	42693
30	81910	00774	04190	33528	18807	64371	78867	49337	69665	38686
31	07273	14043	80432	33541	95806	29973	20871	49814	61299	06543
32	44744	75503	30140	01920	66214	43834	01562	22678	78464	32970
33	59049	26420	56665	02075	73873	37727	51442	25202	49659	39004
34	91194	35877	82939	56681	26239	43530	29668	97729	35020	00290
35	63395	59659	54752	83240	40201	75321	06777	97949	29375	45443
36	96776	91429	80951	20937	51148	22800	93427	42432	39814	03188
37	65703	22165	91226	96651	77591	39259	83294	52333	92003	55615
38	33982	14156	05501	02125	42252	45740	74129	89349	95643	91532
39	62800	01895	26926	98283	19411	58220	11773	07989	25152	65949
40	01650	24083	23285	11298	14134	36497	27930	00930	95384	94789
41	18034	04957	47794	60553	05921	40798	21115	69277	79212	18303
42	33755	31725	34521	76179	50793	72683	54808	85102	54916	20944
43	68668	55013	52031	86387	55883	96150	63309	91525	74170	20909
44	85164	40008	83167	27823	21266	65425	90630	21315	21833	94512
45	35409	43191	80299	29665	28438	01645	03950	82683	14189	47057

Anexo B: Distribuição Normal Padronizada (Z)

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014



Anexo C: Quantis da Distribuição t de Student

	0.400	0.250	0.100	0.050	0.025	0.010	0.001	α
ν	0.600	0.750	0.900	0.950	0.975	0.990	0.999	$1-\alpha$
1	0.3249	1.0000	3.0777	6.3138	12.7062	31.8205	31.8205	
2	0.2887	0.8165	1.8856	2.9200	4.3027	6.9646	6.9646	
3	0.2767	0.7649	1.6377	2.3534	3.1824	4.5407	4.5407	
4	0.2707	0.7407	1.5332	2.1318	2.7764	3.7469	3.7469	
5	0.2672	0.7267	1.4759	2.0150	2.5706	3.3649	3.3649	
6	0.2648	0.7176	1.4398	1.9432	2.4469	3.1427	3.1427	
7	0.2632	0.7111	1.4149	1.8946	2.3646	2.9980	2.9980	
8	0.2619	0.7064	1.3968	1.8595	2.3060	2.8965	2.8965	
9	0.2610	0.7027	1.3830	1.8331	2.2622	2.8214	2.8214	
10	0.2602	0.6998	1.3722	1.8125	2.2281	2.7638	2.7638	
11	0.2596	0.6974	1.3634	1.7959	2.2010	2.7181	2.7181	
12	0.2590	0.6955	1.3562	1.7823	2.1788	2.6810	2.6810	
13	0.2586	0.6938	1.3502	1.7709	2.1604	2.6503	2.6503	
14	0.2582	0.6924	1.3450	1.7613	2.1448	2.6245	2.6245	
15	0.2579	0.6912	1.3406	1.7531	2.1314	2.6025	2.6025	
16	0.2576	0.6901	1.3368	1.7459	2.1199	2.5835	2.5835	
17	0.2573	0.6892	1.3334	1.7396	2.1098	2.5669	2.5669	
18	0.2571	0.6884	1.3304	1.7341	2.1009	2.5524	2.5524	
19	0.2569	0.6876	1.3277	1.7291	2.0930	2.5395	2.5395	
20	0.2567	0.6870	1.3253	1.7247	2.0860	2.5280	2.5280	
21	0.2566	0.6864	1.3232	1.7207	2.0796	2.5176	2.5176	
22	0.2564	0.6858	1.3212	1.7171	2.0739	2.5083	2.5083	
23	0.2563	0.6853	1.3195	1.7139	2.0687	2.4999	2.4999	
24	0.2562	0.6848	1.3178	1.7109	2.0639	2.4922	2.4922	
25	0.2561	0.6844	1.3163	1.7081	2.0595	2.4851	2.4851	
26	0.2560	0.6840	1.3150	1.7056	2.0555	2.4786	2.4786	
27	0.2559	0.6837	1.3137	1.7033	2.0518	2.4727	2.4727	
28	0.2558	0.6834	1.3125	1.7011	2.0484	2.4671	2.4671	
29	0.2557	0.6830	1.3114	1.6991	2.0452	2.4620	2.4620	
30	0.2556	0.6828	1.3104	1.6973	2.0423	2.4573	2.4573	
40	0.2550	0.6807	1.3031	1.6839	2.0211	2.4233	2.4233	
45	0.2549	0.6800	1.3006	1.6794	2.0141	2.4121	2.4121	
50	0.2547	0.6794	1.2987	1.6759	2.0086	2.4033	2.4033	
55	0.2546	0.6790	1.2971	1.6730	2.0040	2.3961	2.3961	
60	0.2545	0.6786	1.2958	1.6706	2.0003	2.3901	2.3901	
65	0.2544	0.6783	1.2947	1.6686	1.9971	2.3851	2.3851	
70	0.2543	0.6780	1.2938	1.6669	1.9944	2.3808	2.3808	
75	0.2542	0.6778	1.2929	1.6654	1.9921	2.3771	2.3771	
80	0.2542	0.6776	1.2922	1.6641	1.9901	2.3739	2.3739	
85	0.2541	0.6774	1.2916	1.6630	1.9883	2.3710	2.3710	
90	0.2541	0.6772	1.2910	1.6620	1.9867	2.3685	2.3685	
95	0.2541	0.6771	1.2905	1.6611	1.9853	2.3662	2.3662	
100	0.2540	0.6770	1.2901	1.6602	1.9840	2.3642	2.3642	
150	0.2538	0.6761	1.2872	1.6551	1.9759	2.3515	2.3515	
200	0.2537	0.6757	1.2858	1.6525	1.9719	2.3451	2.3451	
500	0.2535	0.6750	1.2832	1.6479	1.9647	2.3338	2.3338	
∞	0.2533	0.6745	1.2816	1.6449	1.9600	2.3263	2.3263	

Anexo D: Quantil da Distribuição F

QUANTIL 95%

v_1	v_2									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	161.448	18.513	10.128	7.709	6.608	5.987	5.591	5.318	5.117	4.965
2	199.500	19.000	9.552	6.944	5.786	5.143	4.737	4.459	4.256	4.103
3	215.707	19.164	9.277	6.591	5.409	4.757	4.347	4.066	3.863	3.708
4	224.583	19.247	9.117	6.388	5.192	4.534	4.120	3.838	3.633	3.478
5	230.162	19.296	9.013	6.256	5.050	4.387	3.972	3.687	3.482	3.326
6	233.986	19.330	8.941	6.163	4.950	4.284	3.866	3.581	3.374	3.217
7	236.768	19.353	8.887	6.094	4.876	4.207	3.787	3.500	3.293	3.135
8	238.883	19.371	8.845	6.041	4.818	4.147	3.726	3.438	3.230	3.072
9	240.543	19.385	8.812	5.999	4.772	4.099	3.677	3.388	3.179	3.020
10	241.882	19.396	8.786	5.964	4.735	4.060	3.637	3.347	3.137	2.978
20	248.013	19.446	8.660	5.803	4.558	3.874	3.445	3.150	2.936	2.774
30	250.095	19.462	8.617	5.746	4.496	3.808	3.376	3.079	2.864	2.700
40	251.143	19.471	8.594	5.717	4.464	3.774	3.340	3.043	2.826	2.661
50	251.774	19.476	8.581	5.699	4.444	3.754	3.319	3.020	2.803	2.637
60	252.196	19.479	8.572	5.688	4.431	3.740	3.304	3.005	2.787	2.621
70	252.497	19.481	8.566	5.679	4.422	3.730	3.294	2.994	2.776	2.610
80	252.724	19.483	8.561	5.673	4.415	3.722	3.286	2.986	2.768	2.601
90	252.900	19.485	8.557	5.668	4.409	3.716	3.280	2.980	2.761	2.594
100	253.041	19.486	8.554	5.664	4.405	3.712	3.275	2.975	2.756	2.588
∞	254.314	19.496	8.526	5.628	4.365	3.669	3.230	2.928	2.707	2.538

v_1	v_2									
	20	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**∞**
1	4.351	4.171	4.085	4.034	4.001	3.978	3.960	3.947	3.936	3.841
2	3.493	3.316	3.232	3.183	3.150	3.128	3.111	3.098	3.087	2.996
3	3.098	2.922	2.839	2.790	2.758	2.736	2.719	2.706	2.696	2.605
4	2.866	2.690	2.606	2.557	2.525	2.503	2.486	2.473	2.463	2.372
5	2.711	2.534	2.449	2.400	2.368	2.346	2.329	2.316	2.305	2.214
6	2.599	2.421	2.336	2.286	2.254	2.231	2.214	2.201	2.191	2.099
7	2.514	2.334	2.249	2.199	2.167	2.143	2.126	2.113	2.103	2.010
8	2.447	2.266	2.180	2.130	2.097	2.074	2.056	2.043	2.032	1.938
9	2.393	2.211	2.124	2.073	2.040	2.017	1.999	1.986	1.975	1.880
10	2.348	2.165	2.077	2.026	1.993	1.969	1.951	1.938	1.927	1.831
20	2.124	1.932	1.839	1.784	1.748	1.722	1.703	1.688	1.676	1.571
30	2.039	1.841	1.744	1.687	1.649	1.622	1.602	1.586	1.573	1.459
40	1.994	1.792	1.693	1.634	1.594	1.566	1.545	1.528	1.515	1.394
50	1.966	1.761	1.660	1.599	1.559	1.530	1.508	1.491	1.477	1.350
60	1.946	1.740	1.637	1.576	1.534	1.505	1.482	1.465	1.450	1.318
70	1.932	1.724	1.621	1.558	1.516	1.486	1.463	1.445	1.430	1.293
80	1.922	1.712	1.608	1.544	1.502	1.471	1.448	1.429	1.415	1.274
90	1.913	1.703	1.597	1.534	1.491	1.459	1.436	1.417	1.402	1.257
100	1.907	1.695	1.589	1.525	1.481	1.450	1.426	1.407	1.392	1.243
∞	1.843	1.622	1.509	1.438	1.389	1.353	1.325	1.302	1.283	1.003

QUANTIL 97.5%

ν_1 \ ν_2	1	2	3	4	5	6	7	8	9	10
1	647.789	38.506	17.443	12.218	10.007	8.813	8.073	7.571	7.209	6.937
2	799.500	39.000	16.044	10.649	8.434	7.260	6.542	6.059	5.715	5.456
3	864.163	39.165	15.439	9.979	7.764	6.599	5.890	5.416	5.078	4.826
4	899.583	39.248	15.101	9.605	7.388	6.227	5.523	5.053	4.718	4.468
5	921.848	39.298	14.885	9.364	7.146	5.988	5.285	4.817	4.484	4.236
6	937.111	39.331	14.735	9.197	6.978	5.820	5.119	4.652	4.320	4.072
7	948.217	39.355	14.624	9.074	6.853	5.695	4.995	4.529	4.197	3.950
8	956.656	39.373	14.540	8.980	6.757	5.600	4.899	4.433	4.102	3.855
9	963.285	39.387	14.473	8.905	6.681	5.523	4.823	4.357	4.026	3.779
10	968.627	39.398	14.419	8.844	6.619	5.461	4.761	4.295	3.964	3.717
20	993.103	39.448	14.167	8.560	6.329	5.168	4.467	3.999	3.667	3.419
30	1001.414	39.465	14.081	8.461	6.227	5.065	4.362	3.894	3.560	3.311
40	1005.598	39.473	14.037	8.411	6.175	5.012	4.309	3.840	3.505	3.255
50	1008.117	39.478	14.010	8.381	6.144	4.980	4.276	3.807	3.472	3.221
60	1009.800	39.481	13.992	8.360	6.123	4.959	4.254	3.784	3.449	3.198
70	1011.004	39.484	13.979	8.346	6.107	4.943	4.239	3.768	3.433	3.182
80	1011.908	39.485	13.970	8.335	6.096	4.932	4.227	3.756	3.421	3.169
90	1012.612	39.487	13.962	8.326	6.087	4.923	4.218	3.747	3.411	3.160
100	1013.175	39.488	13.956	8.319	6.080	4.915	4.210	3.739	3.403	3.152
∞	1018.258	39.498	13.902	8.257	6.015	4.849	4.142	3.670	3.333	3.080

ν_1 \ ν_2	20	30	40	50	60	70	80	90	100	∞
1	5.871	5.568	5.424	5.340	5.286	5.247	5.218	5.196	5.179	5.024
2	4.461	4.182	4.051	3.975	3.925	3.890	3.864	3.844	3.828	3.689
3	3.859	3.589	3.463	3.390	3.343	3.309	3.284	3.265	3.250	3.116
4	3.515	3.250	3.126	3.054	3.008	2.975	2.950	2.932	2.917	2.786
5	3.289	3.026	2.904	2.833	2.786	2.754	2.730	2.711	2.696	2.567
6	3.128	2.867	2.744	2.674	2.627	2.595	2.571	2.552	2.537	2.408
7	3.007	2.746	2.624	2.553	2.507	2.474	2.450	2.432	2.417	2.288
8	2.913	2.651	2.529	2.458	2.412	2.379	2.355	2.336	2.321	2.192
9	2.837	2.575	2.452	2.381	2.334	2.302	2.277	2.259	2.244	2.114
10	2.774	2.511	2.388	2.317	2.270	2.237	2.213	2.194	2.179	2.048
20	2.464	2.195	2.068	1.993	1.944	1.910	1.884	1.864	1.849	1.708
30	2.349	2.074	1.943	1.866	1.815	1.779	1.752	1.731	1.715	1.566
40	2.287	2.009	1.875	1.796	1.744	1.707	1.679	1.657	1.640	1.484
50	2.249	1.968	1.832	1.752	1.699	1.660	1.632	1.610	1.592	1.428
60	2.223	1.940	1.803	1.721	1.667	1.628	1.599	1.576	1.558	1.388
70	2.205	1.920	1.781	1.698	1.643	1.604	1.574	1.551	1.532	1.357
80	2.190	1.904	1.764	1.681	1.625	1.585	1.555	1.531	1.512	1.333
90	2.179	1.892	1.751	1.667	1.611	1.570	1.540	1.516	1.496	1.313
100	2.170	1.882	1.741	1.656	1.599	1.558	1.527	1.503	1.483	1.296
∞	2.085	1.787	1.637	1.545	1.482	1.436	1.400	1.371	1.347	1.004

Bibliografia

(Em Reserva na Biblioteca "Helládio do Amaral Mello")

[AZEVEDO & CAMPOS, 1987]
Azevedo, A.G. de; Campos, P.H.B. de 1987 ***Estatística Básica.*** Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda.

[BEIGUELMAN, 1994]
Beiguelman, B. 1994 ***Curso Prático de Bioestatística.*** (3a. edição). Ribeirão Preto: Sociedade Brasileira de Genética.

[IEMMA, 1992]
Iemma, A.F. 1977 ***Estatística Descritiva***. Piracicaba: φσρ Publicações.

[MENDENHALL, 1985]
Mendenhall, W. 1984 ***Probabilidade e Estatística*** (vol. 2). Rio de Janeiro: Editora Campus.

[OLIVEIRA ,1977]
Oliveira Costa, Neto, P.L. de 1977 ***Estatística***. São Paulo: Edgard Blücher.

[ROCHA, 1975]
Rocha, M.V. da 1975 ***Curso de Estatística***. Rio de Janeiro: Fundação Instituto Brasileiro de Geografia e Estatística..

[WALLIS & ROBERTS, 1964]
Wallis, W.A.; Roberts, H.V. 1964 ***Curso de Estatística***. Rio de Janeiro: Editora Fundo Cultural.

Respostas de Exercícios Selecionados

2.1 (a) nominal; (b) de razão; (c) nominal; (d) de razão; (e) nominal; (f) nominal; (g) de razão; (h) nominal; (i) ordinal; (j) de razão; (k) ordinal; (l) ordinal; (m) nominal.
2.2 $H \in [0,100]$; $V \in [0,1000]$; $X \in \{0,1,2,\ldots,\infty\}$; $I \in \{0,2p\}$; $Y \in \{0,1,2,\ldots,\infty\}$; $S \in [0,1]$; $A \in [0,1]$; $E \in [0,1]$; $C \in [0,30]$; $D_1 \in [0,360]$; $D_2 \in [0,100]$; $F \in [0,360]$

7.3 (a) S={ (A,A); (A,V); (A,R); (V,A); (V,V); (V,R); (R,A); (R,V); (R,R) }
(b) S={ (A,A); (A,V); (A,R); (V,V); (V,R); (R,R) } (c) S={ (A,V); (A,R); (V,A); (V,R); (R,A); (R,V) }
(d) S={ (A,V); (A,R); (V,R) }
7.4 (a) S={III; IIF; IFI; FII; IFF; FIF; FFI; FFF} (b) A={IIF; IFI; FII; IFF; FIF; FFI; FFF} (c) B={IFF; FIF; FFI}
7.6 (a) $(5/6)^2 = 25/36$ (b) 10/36 (c) 1/36 (d) 11/36 (e) 1 - P(A) = P(D) = 11/36
7.7 (a) disjuntos (b) disjuntos (c) C é subconjunto de F, ou F implica C (d) disjuntos (e) idênticos (f) disjuntos
7.8 (a) 410/830 (b) 534/830 (c) 189/830 (d) 225/830 (e) 637/830 (f) 379/830 (g) 193/830
7.9 (a) Lote 1 ≤ 0.16; Lote 2 ≤ 0.6; Lote 3 ≤ 0.22 (desigualdade de Boole) (b) Lote 2, menor soma das probabilidades (desigualdade de Boole)
7.10 (a) 27/59 (b) 42/59 (c) 22/27 (d) 5/27 (e) 22/42=11/21 (f) 20/42=10/21
7.11 (a) 0.6550 (b) 0.7143 (c) 1 **7.12** (a) 0.0240 (2.4%) (b) 0.2083 (c) 0.7917 (d) 0.0461 (e) 0.9539
7.13 (a) 0.52 (b) 0.03 (c) 0.27 (d) 0.49 **7.14** 0.66 **7.15** $1 - (5/6)^{20}$ **7.16** $1 - (0.975)^{30}$
7.17 $1-\left[\frac{365}{365}\bullet\frac{364}{365}\bullet\frac{363}{365}\bullet\frac{362}{365}\bullet\cdots\bullet\frac{(365-23)}{365}\right]=1-\left[\frac{P_{365}^{23}}{365^{23}}\right]=0.5073$ **7.18** 9450 **7.19** $(0.99)^{15}$
7.20 (a) $(0.95)^{10}$ (b) $1 - (0.99)^{10}$ (c) $(0.05)^{10}$