CE008 - Introdução à Bioestatística - Primeiro Semestre 2016

Silvia Shimakura
Silvia.Shimakura@ufpr.br
LEG-UFPR

O objetivo desta disciplina é conhecer os conceitos fundamentais de Estatística com aplicações em saúde.

Aulas teóricas: Quinta-feira (Turmas B e D) 16:30-18:30 e Sexta-feira (Turmas A e C) 10:30-12:30, Setor de Ciências da Saúde, 2o. andar.

Avaliação: Três avaliações de igual peso e o exame final.
Avaliação 1: 31/03 (Turmas B e D) e 01/04 (Turmas A e C)
Avaliação 2: 19/05 (Turmas B e D) e 20/05 (Turmas A e C)
Avaliação 3: 02/06, 09/06, 16/06, 23/06 (Turmas B e D) e 03/06, 10/06, 17/06, 24/06 (Turmas A e C): Seminários em grupo
Exame Final: 07/07 (Turmas B e D) e 08/07 (Turmas A e C)

Programa computacional utilizado: Utilizado para fins didáticos. Sistema de análise estatística R que é gratuito e de código aberto.

Manual da biblioteca Rcmdr - interface gráfica desenvolvida para o R.

Bibliografia:
1. Soares, J. F., Siqueira, A. L. (2002) Introducão à Estatística Médica. COOPMED. ISBN: 85-85002-55-7.
2. Reis, E.A.; Reis, I.A. (2001). Análise Descritiva de Dados - Tabelas e Gráficos. Relatório Técnico RTE-04/2001, Depto Estatística-UFMG.
3. Colton, T. (1974). *Statistics in Medicine*. Little, Brown and Company.

Material do curso: Baseado no livro: Introdução à estatística médica (Soares & Siqueira, 2002). As aulas serão dadas no estilo tutorial e estão disponíveis para download e/ou para acesso aqui.

Acesso à página wiki da disciplina.

Tabelas: Tabelas das distribuições usadas neste curso.

Tabela da distribuição normal.

1 Conteúdo

1. **Introdução:** O papel da Estatística na Ciência.
2. **Estatísticas Descritivas:** Sumário de dados, gráfico de barras, gráfico de setores, histogramas, ramo-e-folhas, mediana, moda, desvio padrão, quartis.
3. **Populaçoes e amostras:** Uso de uma amostra para aprender sobre uma população
4. **Intervalos de confiança:** Estimação de parâmetros populacionais a partir de amostras
5. **Testes de hipóteses:** Conceitos e aplicações de testes para uma amostra
6. **Comparação de dois grupos:** Comparação de médias e de proporções

2 Introdução

2.1 O que é Estatística?

Estatística é um conjunto de métodos usados para analisar quantitativamente conjuntos de dados. A Estatística pode ser aplicada em praticamente todas as áreas do conhecimento humano e em algumas áreas recebe um nome especial. Este é o caso da Bioestatística, que trata de aplicações da Estatística em Ciências Biológicas e da Saúde.

A palavra "Estatística" tem *pelo menos* três significados:

1. coleção de informações numéricas ou *dados*,
2. medidas resultantes de um conjunto de dados, como por exemplo médias, proporções e taxas,
3. métodos usados na coleta e interpretação de dados.

Razões para estudar Estatística?

- A disponibilidade de parelhos modernos, muitos dos quais acoplados a computadores, permitem a quantificação de muitos fenômenos. A massa de *dados* gerada precisa ser analisada adequadamente.
- Na Ciência, são realizados estudos *experimentais* ou *observacionais*, em que o interesse é comparar grupos/tratamentos ou ainda determinar fatores prognósticos/risco importantes.
- O material biológico estudado é sempre uma *amostra* e o objetivo final é tirar conclusões sobre toda a *população* de interesse com base na amostra.

Em geral, a disciplina de estatística refere-se a métodos para coleta e descrição dos dados, e então a verificação da força da evidência nos dados pró ou contra certas idéias científicas. A presença de uma *variação* não previsível nos dados faz disso uma tarefa pouco trivial.

2.2 Variação Amostral

Alguns exemplos em que a variação está presente nos dados.

1. **Função pulmonar em pacientes com fibrose cística**

   A pressão inspiratória estática máxima (PImax) é um índice de vigor respiratório muscular. Os seguintes dados mostram a idade (anos) e uma medida de PImax (cm H_2O) de 25 pacientes com fibrose cística.

Sujeito	Idade	PImax
1	7	80
2	7	85
3	8	110
4	8	95
5	8	95
6	9	100
7	11	45
8	12	95
9	12	130
10	13	75
11	13	80
12	14	70
13	14	80
14	15	100
15	16	120
16	17	110
17	17	125
18	17	75
19	17	100
20	19	40
21	19	75
22	20	110
23	23	150
24	23	75
25	23	95

(a) Todos os pacientes com fibrose cística têm o mesmo valor de PImax?

(b) Assumindo que a idade não afeta PImax, qual é um valor de PImax típico para pacientes com fibrose cística?

(c) Quão grande é a variabilidade em torno deste valor típico?

(d) Será que a suposição de que idade não afeta PImax é consistente com os dados?

(e) Se idade na verdade afeta PImax, como você descreveria o valor típico de PImax e variabilidade?

(f) Que tipo de representação gráfica poderia ser utilizada para visualizar adequadamente estes dados?

2. **Conteúdo de gordura e proteína no leite**

   Cientístas mediram o conteúdo de gordura e proteína em amostras de leite de 10 focas cinza.

Foca	Gordura %	Proteína %
1	57.2	10.4
2	58.3	9.4
3	53.9	11.9
4	48.0	12.4
5	57.8	12.1
6	54.1	8.5
7	55.6	10.4
8	49.3	11.6
9	48.8	11.4
10	53.8	10.8

   (a) Os percentuais são exatamente os mesmos de um animal para outro?

   (b) Baseado nesta amostra de 10 focas, os cientístas estimaram o conteúdo de gordura no leite de focas cinza como sendo 53.7%. Se eles agora coletarem mais amostras de leite de outras 10 focas, você esperaria que o novo valor estimado fosse exatamente 53.7%?

   (c) Como o tamanho de amostra influencia sua resposta?

   (d) O que aconteceria se eles tomassem um outro conjunto de amostras das mesmas 10 focas? Você esperaria obter a mesma estimativa neste caso?

   (e) O que aconteceria se uma fração do material coletado inicialmente das 10 focas fosse re-analisado? Você esperaria obter a mesma estimativa neste caso?

Pode-se dizer que cada medida pode ser constituída de três fontes de variação: Variação biológica, variação temporal e variação devido à erros de medida.

3 Estatística Descritiva - Tabelas e Gráficos

Edna A. Reis e Ilka A. Reis
Relatório Técnico RTE-04/2001
Departamento de Estatística-UFMG

A coleta de dados estatísticos tem crescido muito nos últimos anos em todas as áreas de pesquisa, especialmente com o advento dos computadores e surgimento de softwares cada vez mais sofisticados. Ao mesmo tempo, olhar uma extensa listagem de dados coletados não permite obter praticamente nenhuma conclusão, especialmente para grandes conjuntos de dados, com muitas características sendo investigadas.

Utilizamos métodos de Estatística Descritiva para **organizar**, **resumir** e **descrever** os aspectos importantes de um conjunto de características observadas ou comparar tais características entre dois ou mais conjuntos.

As ferramentas descritivas são os muitos tipos de gráficos e tabelas e também medidas de síntese como porcentagens, índices e médias.

Ao se condensar os dados, perde-se informação, pois não se têm as observações originais. Entretanto, esta perda de informação é pequena se comparada ao ganho que se tem com a clareza da interpretação proporcionada.

A descrição dos dados também tem como objetivo **identificar anomalias**, até mesmo resultante do registro incorreto de valores, e dados dispersos, aqueles que não seguem a tendência geral do restante do conjunto. Não só nos artigos técnicos direcionados para pesquisadores, mas também nos artigos de jornais e revistas escritos para o público leigo, é cada vez mais freqüente a utilização destes recursos de descrição para complementar a apresentação de um fato, justificar ou referendar um argumento.

Ao mesmo tempo que o uso das ferramentas estatísticas vem crescendo, aumenta também o **abuso** de tais ferramentas. É muito comum vermos em jornais e revistas, até mesmo em periódicos científicos, gráficos voluntariamente ou intencionalmente enganosos e estatísticas obscuras para justificar argumentos polêmicos.

3.1 Coleta e Armazenamento de Dados

Exemplo Inicial: Ursos Marrons
Pesquisadores do Instituto Amigos do Urso têm estudado o desenvolvimento dos ursos marrons selvagens que vivem em uma certa floresta do Canadá. O objetivo do projeto é estudar algumas características dos ursos, tais como seu peso e altura, ao longo da vida desses animais.

A ficha de coleta de dados, representada na Figura1, mostra as características que serão estudadas na primeira fase do projeto. Na primeira parte do estudo, 97 ursos foram identificados (por nome), pesados e medidos. Os dados foram coletados através do preenchimento da ficha de coleta.

Para que os ursos possam ser identificados, medidos e avaliados, os pesquisadores precisam anestesiá-los. Mesmo assim, medidas como a do peso são difíceis de serem feitas (qual será o

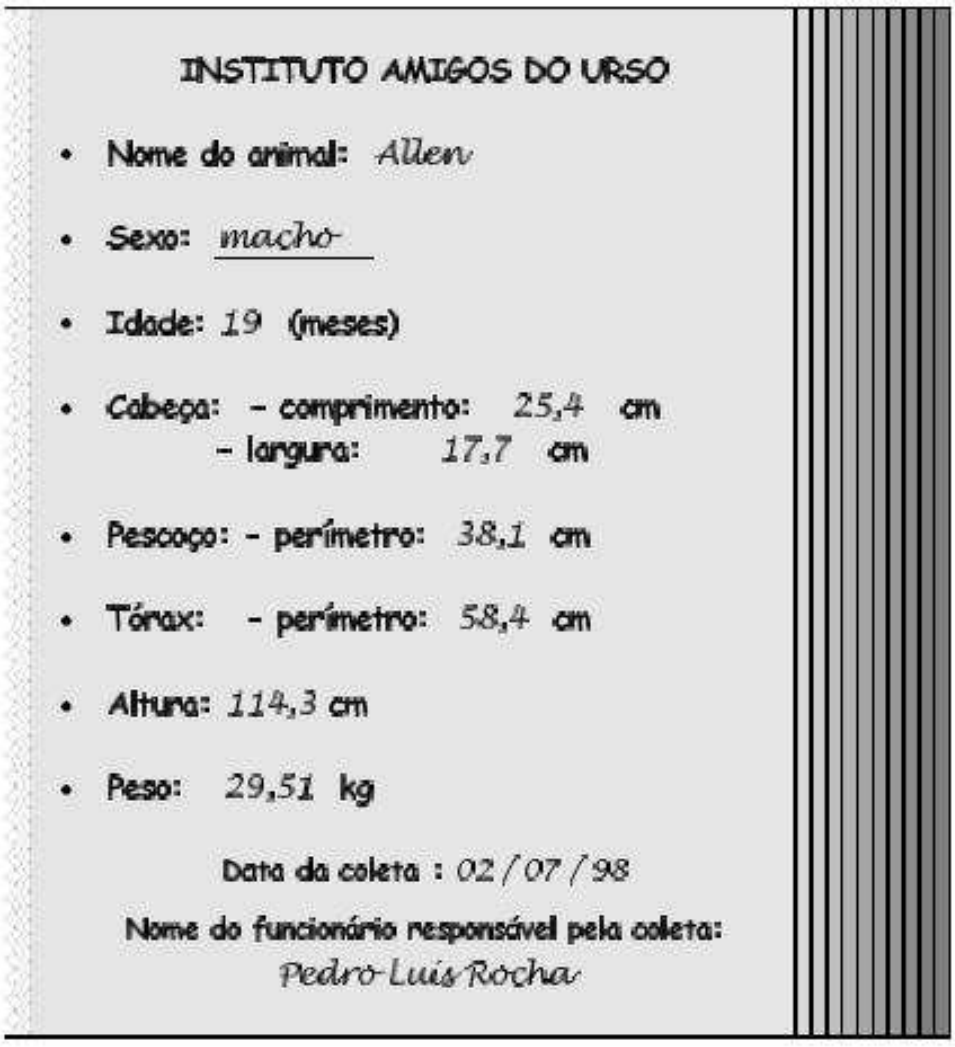


Figura 1: Ficha de coleta de dados dos ursos marrons.

tamanho de uma balança para pesar ursos ?). Desse modo, os pesquisadores gostariam também de encontrar uma maneira de estimar o peso do urso através de uma outra medida mais fácil de se obter, como uma medida de comprimento, por exemplo (altura, circunferência do tórax, etc.). Nesse caso, só seria necessária uma grande fita métrica, o que facilitaria muito a coleta de dados das próximas fases do projeto.

Geralmente, as coletas de dados são feitas através do preenchimento de fichas pelo pesquisador e/ou através de resposta a questionários (o que não foi o caso dos ursos é claro!). Alguns dados são coletados através de medições (altura, peso, pressão sangüínea, etc.), enquanto outros são coletados através de avaliações (sexo, cor, raça, espécie, etc.).

Depois de coletados, os dados devem ser armazenados e sistematizados numa planilha de dados, como mostra a Figura 2. Hoje em dia, essas planilhas são digitais e essa é a maneira de realizar a entrada dos dados num programa de computador.

A planilha de dados é composta por linhas e colunas. Cada linha contém os dados de uma

VARIÁVEIS →

ELEMENTOS ↓	Nome	Mês Obs.	Idade	Sexo	Cabeça Comp.	Cabeça Larg.	Pescoço Peri.	Altura	Tórax Peri.	Peso
1	Allen	jul	19	macho	25,4	12,7	38,1	114,3	58,4	29,5
2	Berta	jul	19	fêmea	27,9	16,5	50,8	120,7	61,0	31,8
3	Clyde	jul	19	macho	27,9	14,0	40,6	134,6	66,0	36,3
4	Doc	jul	55	macho	41,9	22,9	71,1	171,5	114,3	156,2
5	Quincy	set	81	macho	39,4	20,3	78,7	182,9	137,2	188,9
6	Kooch	out	*	macho	40,6	20,3	81,3	195,6	132,1	196,1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
93	Sara	ago	*	fêmea	30,5	12,7	45,7	142,2	82,6	51,8
94	Lou	ago	*	macho	30,5	14,0	38,1	129,5	61,0	37,2
95	Molly	ago	*	fêmea	33,0	15,2	55,9	154,9	101,6	104,4
96	Graham	jul	*	macho	30,5	10,2	44,5	149,9	72,4	58,1
97	Jeffrey	jul	*	macho	34,3	15,2	50,8	157,5	82,6	70,8

Figura 2: Representação parcial da planilha de dados do exemplo dos ursos.

unidade experimental (urso), ou seja de uma ficha de coleta. As características (variáveis) são dispostas em colunas. Assim, a planilha de dados contém um número de linhas igual a número de participantes do estudo e um número de colunas igual ao número de variáveis sendo estudadas.

A planilha de dados dos ursos tem 97 linhas e 10 colunas. Alguns ursos não tiveram sua idade determinada. Esses dados são chamados dados faltantes e é comum representá-los por asteriscos (na verdade, cada software tem sua convenção para representar missing data).

3.2 Tipos de variáveis

Variável é a característica de interesse que é medida em cada elemento da amostra ou população. Como o nome diz, seus valores variam de elemento para elemento. As variáveis podem ter valores numéricos ou não numéricos.

Variáveis podem ser classificadas da seguinte forma:

1. **Variáveis Quantitativas:** são as características que podem ser medidas em uma escala quantitativa, ou seja, apresentam valores numéricos que fazem sentido. Podem ser contínuas ou discretas.

   (a) **Variáveis discretas:** características mensuráveis que podem assumir apenas um número finito ou infinito contável de valores e, assim, somente fazem sentido valores inteiros. Geralmente são o resultado de contagens. Exemplos: número de filhos, número de bactérias por litro de leite, número de cigarros fumados por dia.

   (b) **Variáveis contínuas**, características mensuráveis que assumem valores em uma escala contínua (na reta real), para as quais valores fracionais fazem sentido. Usual-

mente devem ser medidas através de algum instrumento. Exemplos: peso (balança), altura (régua), tempo (relógio), pressão arterial, idade.

2. **Variáveis Qualitativas (ou categóricas)**: são as características que não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos. Podem ser nominais ou ordinais.

   (a) **Variáveis nominais**: não existe ordenação dentre as categorias. Exemplos: sexo, cor dos olhos, fumante/não fumante, doente/sadio.

   (b) **Variáveis ordinais**: existe uma ordenação entre as categorias. Exemplos: escolaridade (1o, 2o, 3o graus), estágio da doença (inicial, intermediário, terminal), mês de observação (janeiro, fevereiro,..., dezembro).

As distinções são menos rígidas do que a descrição acima insinua.

Uma variável originalmente quantitativa pode ser coletada de forma qualitativa.
Por exemplo, a variável idade, medida em anos completos, é quantitativa (contínua); mas, se for informada apenas a faixa etária (0 a 5 anos, 6 a 10 anos, etc...), é qualitativa (ordinal). Outro exemplo é o peso dos lutadores de boxe, uma variável quantitativa (contínua) se trabalhamos com o valor obtido na balança, mas qualitativa (ordinal) se o classificarmos nas categorias do boxe (peso-pena, peso-leve, peso-pesado, etc.).

Outro ponto importante é que *nem sempre uma variável representada por números é quantitativa.*
O número do telefone de uma pessoa, o número da casa, o número de sua identidade. Às vezes o sexo do indivíduo é registrado na planilha de dados como 1 se macho e 2 se fêmea, por exemplo. Isto não significa que a variável sexo passou a ser quantitativa!

Exemplo do ursos marrons (continuação):
No conjunto de dados ursos marrons, são qualitativas as variáveis sexo (nominal) e mês da observação (ordinal); são quantitativas contínuas as demais: idade, comprimento da cabeça, largura da cabeça, perímetro do pescoço, perímetro do tórax, altura e peso.

3.3 Estudando a Distribuição de Freqüências de uma Variável

Como já sabemos, as variáveis de um estudo dividem-se em quatro tipos: qualitativas (nominais e ordinais) e quantitativas (discretas e contínuas). Os dados gerados por esses tipos de variáveis são de naturezas diferentes e devem receber tratamentos diferentes. Portanto, vamos estudar as ferramentas - **tabelas e gráficos** - mais adequados para cada tipo de dados, separadamente.

3.3.1 Variáveis Qualitativas - Nominais e Ordinais

Iniciaremos essa apresentação com os dados de natureza qualitativa, que são os mais fáceis de tratar do ponto de vista da análise descritiva.

No exemplo dos ursos, uma das duas variáveis qualitativas presentes é o **sexo** dos animais.

Para organizar os dados provenientes de uma variável qualitativa, é usual fazer uma tabela de freqüências, como a Tabela 1, onde estão apresentadas as freqüências com que ocorrem cada um dos sexos no total dos 97 ursos observados.

Cada categoria da variável sexo (feminino, masculino) é representada numa linha da tabela. Há uma coluna com as contagens de ursos em cada categoria (freqüência absoluta) e outra com os percentuais que essas contagens representam no total de ursos (freqüência relativa). Esse tipo de tabela representa a distribuição de freqüências dos ursos segundo a variável sexo.

Como a variável sexo é **qualitativa nominal**, isto é, não há uma ordem natural em suas categorias, a ordem das linhas da tabela pode ser qualquer uma.

Tabela 1: Distribuição de freqüências dos ursos segundo sexo.

Sexo	Freqüência Absoluta	Freqüência Relativa (%)
Feminino	35	36,1
Masculino	62	63,9
Total	97	100,0

Quando a variável tabelada for do tipo **qualitativa ordinal**, as linhas da tabela de freqüências devem ser dispostas na ordem existente para as categorias.

A Tabela 2 mostra a distribuição de freqüências dos ursos segundo o mês de observação, que é uma variável qualitativa ordinal. Nesse caso, podemos acrescentar mais duas colunas com as freqüências acumuladas (absoluta e relativa), que mostram, para cada mês, a freqüência de ursos observados até aquele mês. Por exemplo, até o mês de julho, foram observados 31 ursos, o que representa 32,0% do total de ursos estudados.

Tabela 2: Distribuição de freqüências dos ursos segundo mês de observação.

	Freqüências Simples		Freqüências Acumuladas	
			Freqüência	Freqüência
Mês de Observação	Freqüência Absoluta	Freqüência Relativa (%)	Absoluta Acumulada	Relativa Acumulada
Abril	8	8,3	8	8,3
Maio	6	6,2	14	14,5
Junho	6	6,2	20	20,7
Julho	11	11,3	31	32,0
Agosto	23	23,7	54	55,7
Setembro	20	20,6	74	76,3
Outubro	14	14,4	88	90,7
Novembro	9	9,3	97	100,0
Total	97	100,0	—	—

A visualização da distribuição de freqüências de uma variável fica mais fácil se fizermos um gráfico a partir da tabela de freqüências. Existem vários tipos de gráficos, dependendo do tipo de variável a ser representada. Para as variáveis do tipo qualitativas, abordaremos dois tipos de gráficos: **os de setores e os de barras**.

Os gráficos de setores, mais conhecidos como gráficos de pizza ou torta, são construídos dividindo-se um círculo (pizza) em setores (fatias), um para cada categoria, que serão proporcionais à freqüência daquela categoria.

A Figura 3 mostra um gráfico de setores para a variável sexo, construído a partir da Tabela 1. Através desse gráfico, fica mais fácil perceber que os ursos machos são a grande maioria dos ursos estudados. Como esse gráfico contém todas as informações da Tabela 1, pode substituí-la com a vantagem de tornar análise dessa variável mais agradável.

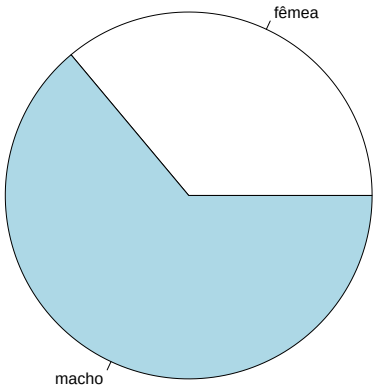


Figura 3: Gráfico de setores para a variável sexo.

As vantagens da representação gráfica das distribuições de freqüências ficam ainda mais evidentes quando há a necessidade de comparar vários grupos com relação à variáveis que possuem muitas categorias, como veremos mais adiante.

Uma alternativa ao gráfico de setores é o gráfico de barras (colunas) como o da Figura 4. Ao invés de dividirmos um círculo, dividimos uma barra. Note que, em ambos os gráficos, as freqüências relativas das categorias devem somar 100%. Aliás, essa é a idéia dos gráficos: mostrar como se dá a divisão (distribuição) do total de elementos (100%) em partes (fatias).

Uma situação diferente ocorre quando desejamos comparar a distribuição de freqüências de uma mesma variável em vários grupos, como por exemplo, a freqüência de ursos marrons em quatro regiões de um país.

Se quisermos usar o gráfico de setores para fazer essa comparação, devemos fazer quatro gráficos, um para cada região, com duas fatias cada um (ursos marrons e ursos não marrons). Uma

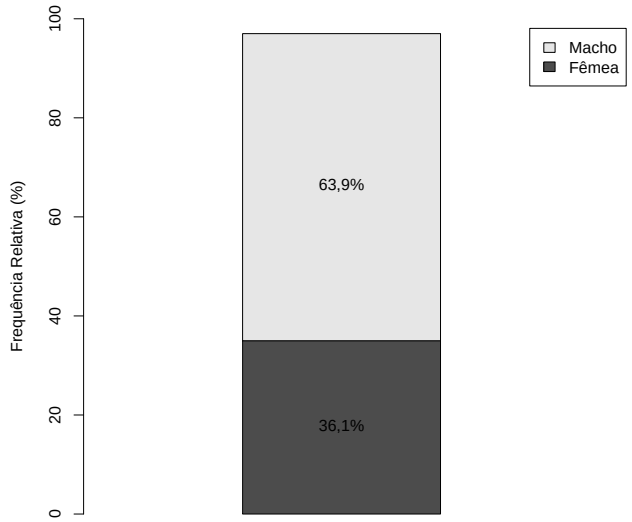


Figura 4: Gráfico de barras para a variável sexo.

alternativa é a construção de um gráfico de barras (horizontal ou vertical) como na Figura 5, com uma barra para cada região representando a freqüência de ursos marrons naquela região. Além de economizar espaço na apresentação, permite que as comparações sejam feitas de maneira mais rápida (tente fazer essa comparação usando quatro pizzas e comprove!!)

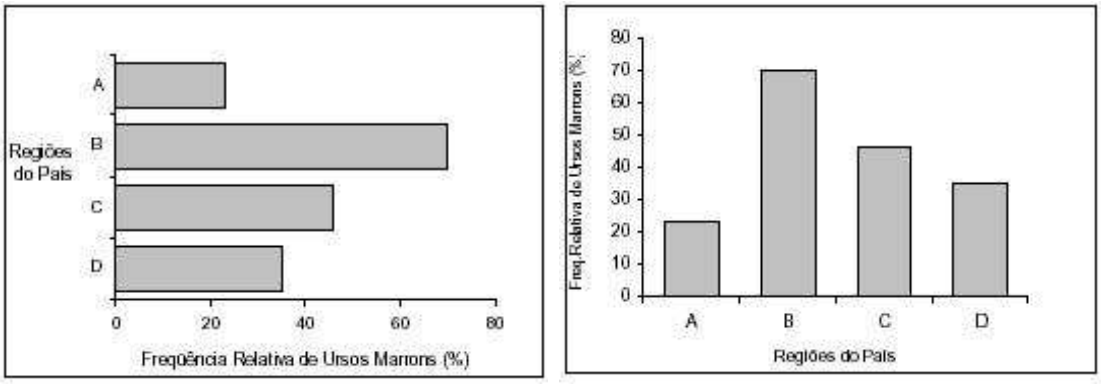


Figura 5: Gráfico de barras horizontais e verticais para a freqüência de ursos marrons em quatro regiões.

A ordem dos grupos pode ser qualquer, ou aquela mais adequada para a presente análise. Freqüentemente, encontramos as barras em ordem decrescente, já antecipando nossa intuição de ordenar os grupos de acordo com sua freqüência para facilitar as comparações. Caso a variável fosse do tipo ordinal, a ordem das barras seria a ordem natural das categorias, como na tabela de freqüências.

A Figura 6 mostra um gráfico de barras que pode ser usado da comparação da distribuição de freqüências de uma mesma variável em vários grupos. É também uma alternativa ao uso de vários gráficos de setores, sendo, na verdade, a junção de três gráficos com os da Figura 4 num só gráfico.

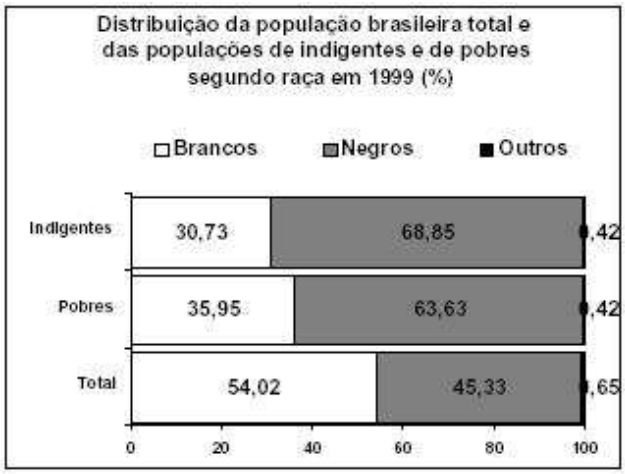


Figura 6: Gráfico de barras para comparação da distribuição de freqüências de uma variável (raça) em vários grupos (indigentes, pobres e população total).

Observação: Este tipo de gráfico só deve ser usado quando não houver muitos grupos a serem comparados e a variável em estudo não tiver muitas categorias (de preferência, só duas). No exemplo da Figura 6, a variável raça tem três categorias, mas uma delas é muito menos freqüente do que as outras duas.

Através desse gráfico, podemos observar que a população brasileira total, em 1999, dividia-se quase que igualmente entre brancos e negros, com uma pequena predominância de brancos. Porém, quando nos restringimos às classes menos favorecidas economicamente, essa situação se inverte, com uma considerável predominância de negros, principalmente na classe da população considerada indigente, indicando que a classe sócio-econômica influencia a distribuição de negros e brancos na população brasileira de 1999.

Freqüentemente, é necessário fazer comparações da distribuição de freqüências de uma variável em vários grupos simultaneamente. Nesse caso, o uso de gráficos bem escolhidos e construídos torna a tarefa muito mais fácil. Na Figura 7, está representada a distribuição de freqüências da reprovação segundo as variáveis sexo do aluno, período e área de estudo.

Analisando os três gráficos da Figura 7, podemos notar que o percentual de reprovação entre os alunos do sexo masculino é sempre maior do que o percentual de reprovação entre os alunos do sexo feminino, em todas as áreas, durante todos os períodos.

A área de ciências exatas é a que possui os maiores percentuais de reprovação, em todos os períodos, nos dois sexos.

Na área de ciências humanas, o percentual de reprovação entre os alunos do sexo masculino cresce com os períodos, enquanto esse percentual entre as alunas se mantém praticamente constante durante os períodos.

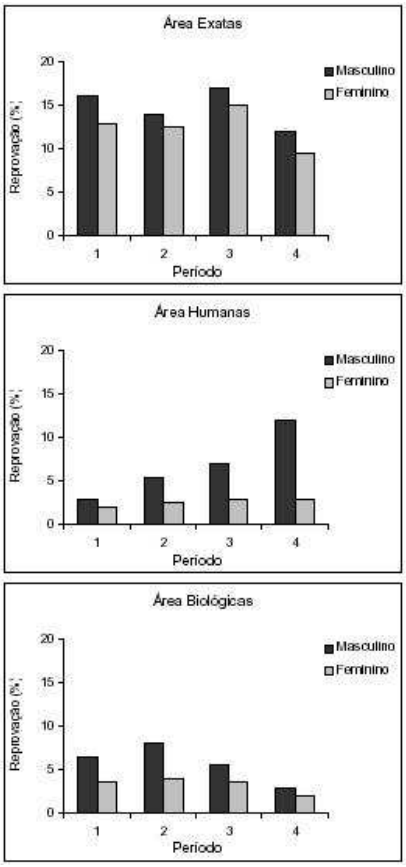


Figura 7: Distribuição de freqüências de reprovação segundo área, período e sexo do aluno. Fonte: A Evasão no Ciclo Básico da UFMG, em Cadernos de Avaliação 3, 2000.

Na área de ciências biológicas, há uma diminuição do percentual de reprovação, a partir do segundo período, entre os alunos do dos sexos, sendo mais acentuado entre os estudantes do sexo masculino.

Chegar às conclusões colocadas acima através de comparação numérica de tabelas de freqüências seria muito mais árduo do que através da comparação visual possibilitada pelo uso dos gráficos. Os gráficos são ferramentas poderosas e devem ser usadas sempre que possível.

É importante observar que a comparação dos três gráficos da Figura 7 só foi possível porque eles usam a mesma escala, tanto no eixo dos períodos (mesma ordem) quanto no eixo dos percentuais de reprovação (mais importante). Essa observação é válida para toda comparação entre gráficos de quaisquer tipo.

3.3.2 Variáveis Quantitativas Discretas

Quando estamos trabalhando com uma variável discreta que assume poucos valores, podemos dar a ela o mesmo tratamento dado às variáveis qualitativas ordinais, assumindo que cada valor é uma classe e que existe uma ordem natural nessas classes.

A Tabela 3 apresenta a distribuição de freqüências do número de filhos por família em uma

localidade, que, nesse caso, assumiu apenas seis valores distintos.

Tabela 3: Distribuição de freqüências do número de filhos por família em uma localidade (25 lares).

Número de filhos	Freqüência Absoluta	Freqüência Relativa (%)	Frequência Relativa Acumulada (%)
0	1	4,0	4,0
1	4	16,0	20,0
2	10	40,0	60,0
3	6	24,0	84,0
4	2	8,0	92,0
5	2	8,0	100,0
Total	25	100	—

Analisando a Tabela 3, podemos perceber que as famílias mais freqüentes são as de dois filhos (40%), seguida pelas famílias de três filhos. Apenas 16% das famílias têm mais de três filhos, mas são ainda mais comuns do que famílias sem filhos.

A Figura 8 mostra a representação gráfica da Tabela 3 no gráfico à esquerda e a distribuição de freqüências do número de filhos por família na localidade B no gráfico à direita. Como o número de famílias estudadas em cada localidade é diferente, a freqüência utilizada em ambos os gráficos foi a relativa (em porcentagem), tornando os dois gráficos comparáveis. Comparando os dois gráficos, notamos que a localidade B tende a ter famílias menos numerosas do que a localidade A. A maior parte das famílias da localidade B (cerca de 70%) têm um ou nenhum filho.

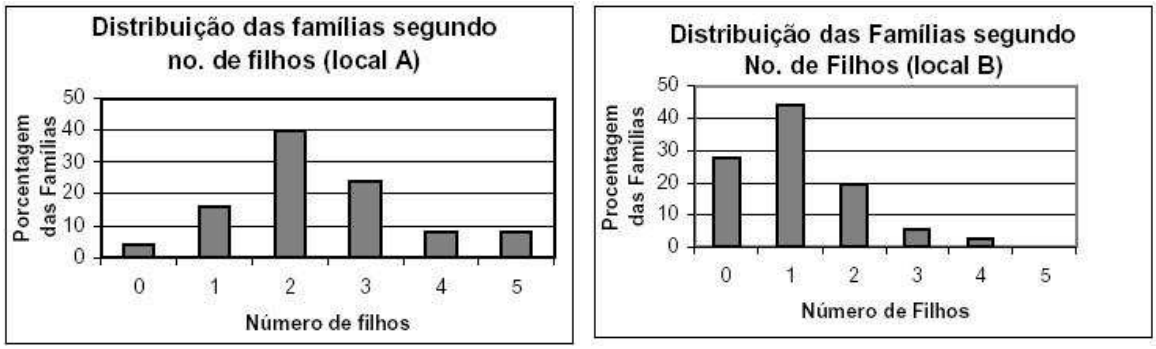


Figura 8: Distribuição de freqüências do número de filhos por família na localidade A (25 lares) e B (36 lares).

Importante: Na comparação da distribuição de freqüências de uma variável entre dois ou mais grupos de tamanhos (número de observações) diferentes, devemos usar as freqüências relativas na construção do histograma. Deve-se, também usar a mesma escala em todos os histogramas, tanto na escala vertical quanto na horizontal.

Quando trabalhamos com uma variável discreta que pode assumir um grande número de valores distintos como, por exemplo, o número de ovos que um inseto põe durante sua vida, a construção da tabela de freqüências e de gráficos considerando cada valor como uma categoria fica inviável. A solução é agrupar os valores em classes ao montar a tabela, como mostra a Tabela 4.

Tabela 4: Distribuição de freqüências do número de ovos postos por 250 insetos.

	Freqüências Simples		Freqüências Acumuladas	
Número de ovos	Freqüência Absoluta	Freqüência Relativa (%)	Freq.Abs. Acumulada	Freq.Rel. Acumulada(%)
10 a 14	4	1,6	4	1,6
15 a 19	30	12,0	34	13,6
20 a 24	97	38,8	131	52,4
25 a 29	77	30,8	208	83,2
30 a 34	33	13,2	241	96,4
35 a 39	7	2,8	248	99,2
40 a 44	2	0,8	250	100,0
Total	250	100	—	—

A Figura 9 mostra o gráfico da distribuição de freqüências do número de ovos postos por 250 insetos ao longo de suas vidas. Podemos perceber que o número de ovos está concentrado em torno de 20 a 24 ovos com um ligeiro deslocamento para os valores maiores.

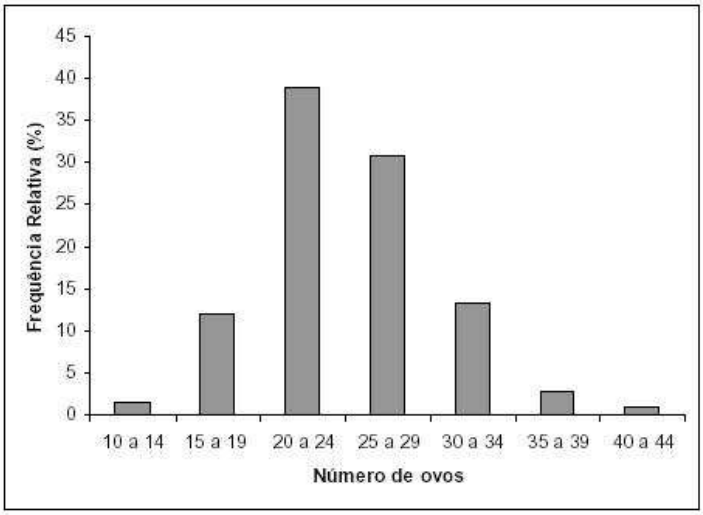


Figura 9: Distribuição de freqüências do número de ovos postos por 250 insetos.

A escolha do número de classes e do tamanho das classes depende da amplitude dos valores a serem representados (no exemplo, de 10 a 44) e da quantidade de observações no conjunto de dados.

Classes muito grandes resumem demais a informação contida nos dados, pois forçam a construção

de poucas classes. No exemplo dos insetos, seria como, por exemplo, construir classes de tamanho 10, o que reduziria para quatro o número de classes (Figura 10).

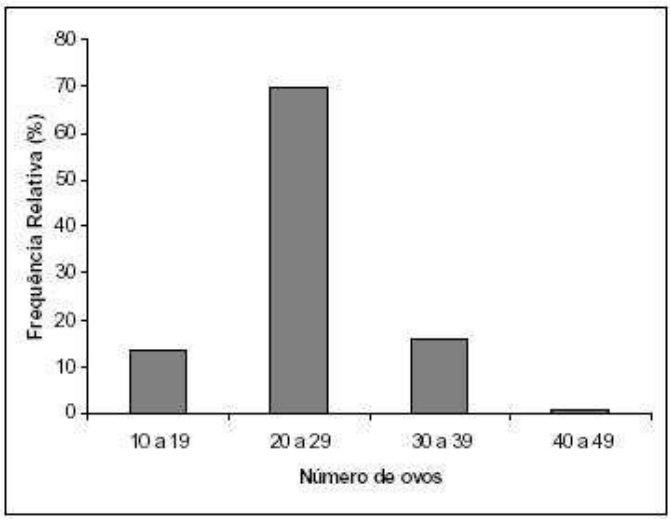


Figura 10: Distribuição de freqüências do número de ovos postos por 250 insetos.(classes de tamanho 10)

Por outro lado, classes muito pequenas nos levaria a construir muitas classes, o que poderia não resumir a informação como gostaríamos. Além disso, para conjuntos de dados pequenos, pode ocorrer classes com muito poucas observações ou mesmo sem observações. Na Figura 11, há classes sem observações, mesmo o conjunto de dados sendo grande.

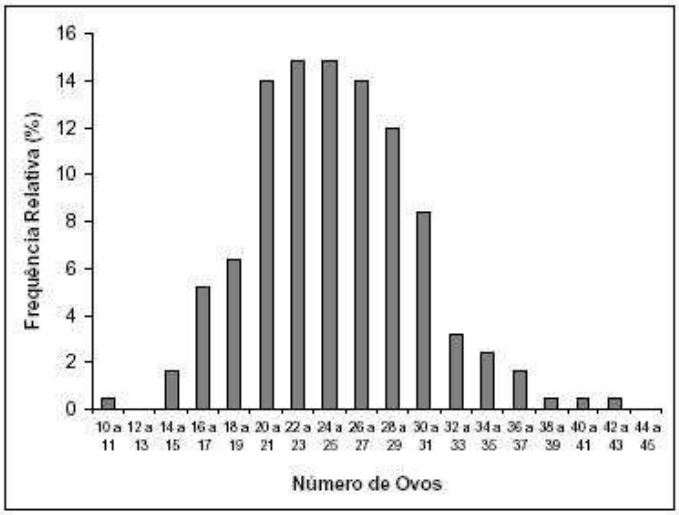


Figura 11: Distribuição de freqüências do número de ovos postos por 250 insetos.(classes de tamanho 2)

Alguns autores recomendam que tabelas de freqüências (e gráficos) possuam de 5 a 15 classes, dependendo do tamanho do conjunto de dados e levando-se em consideração o que foi exposto anteriormente.

Os limites inferiores e superiores de cada classe dependem do tamanho (amplitude) de classe escolhido, que deve ser, na medida do possível, igual para todas as classes. Isso facilita a interpretação da distribuição de freqüências da variável em estudo.

Com o uso do computador na análise estatística de dados, a tarefa de construção de tabelas e gráficos ficou menos trabalhosa e menos dependente de regras rígidas. Se determinado agrupamento de classes não nos pareceu muito bom, podemos construir vários outros quase que instantaneamente e a escolha da melhor representação para a distribuição de freqüências para aquela variável fica muito mais tranqüila.

3.3.3 Variáveis Quantitativas Contínuas

Quando a variável em estudo é do tipo contínua, que assume muitos valores distintos, o agrupamento dos dados em classes será sempre necessário na construção das tabelas de freqüências. A Tabela 5 apresenta a distribuição de freqüências para o peso dos ursos machos.

Tabela 5: Distribuição de freqüências dos ursos machos segundo peso.

Peso (kg)	Freqüência Absoluta	Freqüência Relativa (%)	Freq. Abs. Acumulada	Freq. Rel. Acumulada (%)
$0\| - 25$	3	4,8	3	4,8
$25\| - 50$	11	17,7	14	22,6
$50\| - 75$	15	24,2	29	46,8
$75\| - 100$	11	17,7	40	64,5
$100\| - 125$	3	4,8	43	69,4
$125\| - 150$	4	6,5	47	75,8
$150\| - 175$	8	12,9	55	88,7
$175\| - 200$	5	8,1	60	96,8
$200\| - 225$	1	1,6	61	98,4
$225\| - 250$	1	1,6	62	100,0
Total	62	100,0	-	-

Os limites das classes são representados de modo diferente daquele usado nas tabelas para variáveis discretas: o limite superior de uma classe é igual ao limite inferior da classe seguinte. Mas, afinal, onde ele está incluído?

O símbolo $|-$ resolve essa questão. Na segunda classe ($25| - 50$), por exemplo, estão incluídos todos os ursos com peso de 25,0 a 49,9 kg. Os ursos que porventura pesarem exatos 50,0 kg serão incluídos na classe seguinte. Ou seja, ursos com pesos maiores ou iguais a 25 kg e menores do que 50 kg.

A construção das classes da tabela de freqüências é feita de modo a facilitar a interpretação da distribuição de freqüências, como discutido anteriormente. Geralmente, usamos tamanhos e limites de classe múltiplos de 5 ou 10. Isso ocorre porque estamos acostumados a pensar no nosso sistema numérico, que é o decimal. Porém, nada nos impede de construirmos classes de outros

tamanhos (inteiros ou fracionários) desde que isso facilite nossa visualização e interpretação da distribuição de freqüências da variável em estudo.

A representação gráfica da distribuição de freqüências de uma variável contínua é feita através de um gráfico chamado histograma, mostrado na Figura 12. O histograma nada mais é do que o gráfico de barras verticais, porém construído com as barras unidas, devido ao caráter contínuo dos valores da variável.

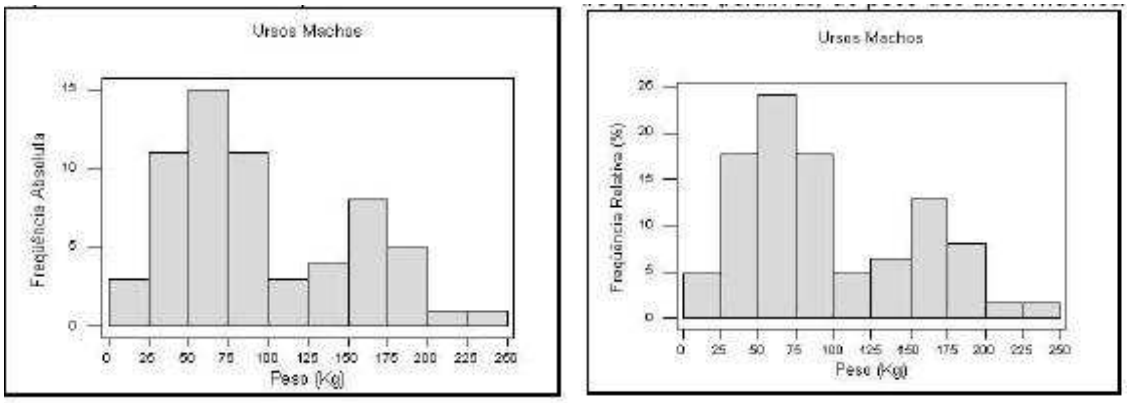


Figura 12: Histograma para a distribuição de frequências (absolutas e relativas) de pesos de ursos machos

Os histogramas da Figura 12 têm a mesma forma, apesar de serem construídos usando as freqüências absolutas e relativas, respectivamente. O objetivo dessas figuras é mostrar que a escolha do tipo de freqüência a ser usada não muda a forma da distribuição. Entretanto, o uso da freqüência relativa torna o histograma comparável a outros histogramas, mesmo que os conjuntos de dados tenham tamanhos diferentes (desde a mesma escala seja usada!)

Analisando o histograma para o peso dos ursos machos, podemos perceber que há dois grupos de ursos: os mais leves, com pesos em torno de 50 a 75 Kg, e os mais pesados, com pesos em torno de 150 a 175 Kg. Essa divisão pode ser devida a uma outra característica dos ursos, como idades ou hábitos alimentares diferentes, por exemplo.

A Tabela 6 apresenta a distribuição de freqüências para o peso dos ursos fêmeas, representada graficamente pelo histograma à esquerda na Figura 13. Apesar de não haver, neste conjunto de dados, fêmeas com peso maior de que 175 Kg, as três últimas classes foram mantidas para que pudessemos comparar machos e fêmeas quanto ao peso.

A Figura 13 também mostra o histograma para o peso dos ursos machos (à direita). Note que ele tem a mesma forma dos histogramas da Figura 12, porém com as barras mais achatadas, devido à mudança de escala no eixo vertical para torná-lo comparável ao histograma das fêmeas.

Comparando as distribuições dos pesos dos ursos machos e fêmeas, podemos concluir que as fêmeas são, em geral, menos pesadas do que os machos, distribuindo-se quase simetricamente em torno da classe de 50 a 75 Kg . O peso das fêmeas é mais homogêneo (valores mais próximos entre si) do que o peso dos ursos machos.

Tabela 6: Distribuição de freqüências dos ursos fêmeas segundo peso.

Peso (kg)	Freqüência Absoluta	Freqüência Relativa (%)	Freq. Abs. Acumulada	Freq. Rel. Acumulada (%)
$0\| - 25$	3	8,6	3	8,6
$25\| - 50$	5	14,3	8	22,9
$50\| - 75$	18	51,4	26	74,3
$75\| - 100$	5	14,3	31	88,6
$100\| - 125$	2	5,7	33	94,3
$125\| - 150$	1	2,9	34	97,1
$150\| - 175$	1	2,9	35	100,0
$175\| - 200$	0	0	35	100,0
$200\| - 225$	0	0	35	100,0
$225\| - 250$	0	0	35	100,0
Total	35	100,0	-	-

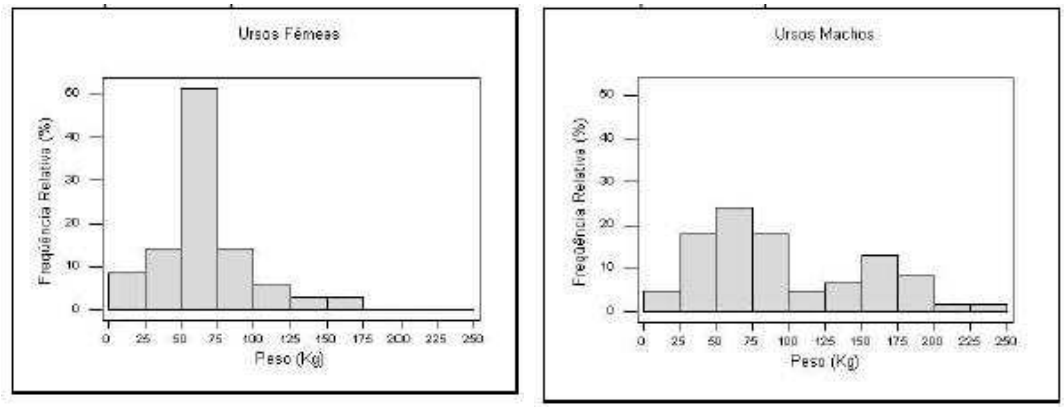


Figura 13: Histograma para a distribuição de frequências de pesos de ursos fêmeas (esquerda) e machos (direita)

Muitas vezes, a análise da distribuição de freqüências acumuladas é mais interessante do que a de freqüências simples, representada pelo histograma. O gráfico usado na representação gráfica da distribuição de freqüências acumuladas de uma variável contínua é a ogiva, apresentada na Figura 14. Para a construção da ogiva, são usadas as freqüências acumuladas (absolutas ou relativas) no eixo vertical e os limites superiores de classe no eixo horizontal.

O primeiro ponto da ogiva é formado pelo limite inferior da primeira classe e o valor zero, indicando que abaixo do limite inferior da primeira classe não existem observações. Daí por diante, são usados os limites superiores das classes e suas respectivas freqüências acumuladas, até a última classe, que acumula todas as observações. Assim, uma ogiva deve começar no valor zero e, se for construída com as freqüências relativas acumuladas, terminar com o valor 100

A ogiva permite que sejam respondidas perguntas do tipo:

a) Qual o percentual de ursos têm peso de até 125 Kg?

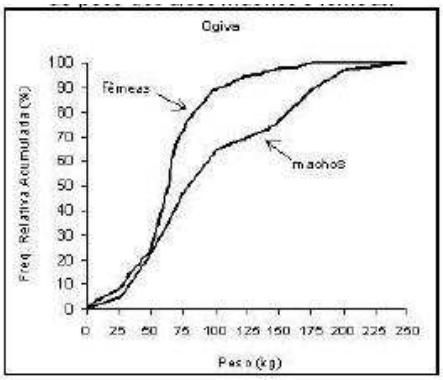


Figura 14: Ogivas para a distribuição de frequências de pesos de ursos machos e fêmeas

Na Figura 15(a), traçamos uma linha vertical partindo do ponto 125 kg até cruzar com cada ogiva (fêmeas e machos). A partir deste ponto de cruzamento, traçamos uma linha horizontal até o eixo das freqüências acumuladas, encontrando o valor de 70% para os machos e 95% para as fêmeas.

Assim, 95% das fêmeas têm até 125 kg, enquanto 70% dos machos têm até 125 kg. É o mesmo que dizer que apenas 5% das fêmeas pesam mais que 125 kg, enquanto 30% dos machos pesam mais que 125 kg.

b) Qual o valor do peso que deixa abaixo (e acima) dele 50% dos ursos?

Na Figura 15(b), traçamos uma linha horizontal partindo da freqüência acumulada de 50% até encontrar as duas ogivas. A partir destes pontos de encontro, traçamos uma linha vertical até o eixo do valores de peso, encontrando o valor de 80 kg para os machos e 65 kg para as fêmeas.

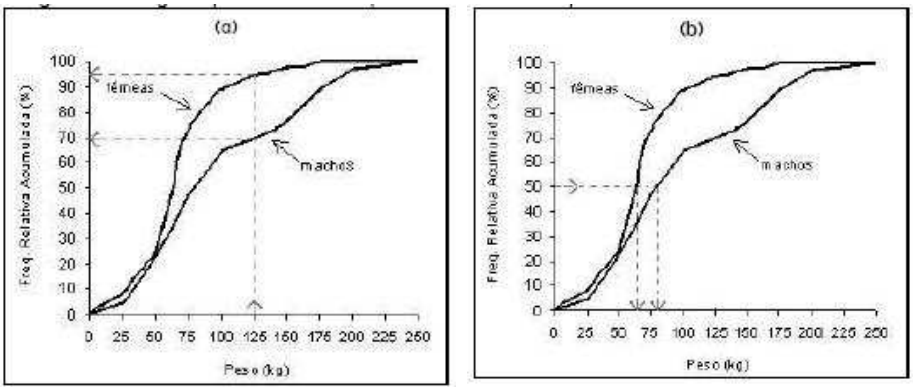


Figura 15: Ogivas para a distribuição de frequências de pesos de ursos machos e fêmeas

Assim, metade dos machos pesam até 80 kg (e metade pesam mais que 80 kg), enquanto metade das fêmeas pesam até 65 kg.

3.3.4 Outros Gráficos para Variáveis Quantitativas

Quando construímos uma tabela de freqüências para uma variável quantitativa utilizando agrupamento de valores em classes, estamos resumindo a informação contida nos dados. Isto é desejável quando o número de dados é grande e sem um algum tipo de resumo ficaria difícil tirar conclusões sobre o comportamento da variável em estudo.

Porém, quando a quantidade de dados disponíveis não é tão grande, o resumo promovido pelo histograma não é aconselhável.

Para os casos em que o número de dados é pequeno, uma alternativa para a visualização da distribuição desses dados são os gráficos denominados diagrama de pontos e diagrama de ramo-e-folhas.

O Diagrama de Pontos

Uma representação alternativa ao histograma para a distribuição de freqüências de uma variável quantitativa é o diagrama de pontos, como aqueles mostrado mostrados na Figura 16.

Neste gráfico, cada ponto representa uma observação com determinado valor da variável. Observações com mesmo valor são representadas com pontos empilhados neste valor.

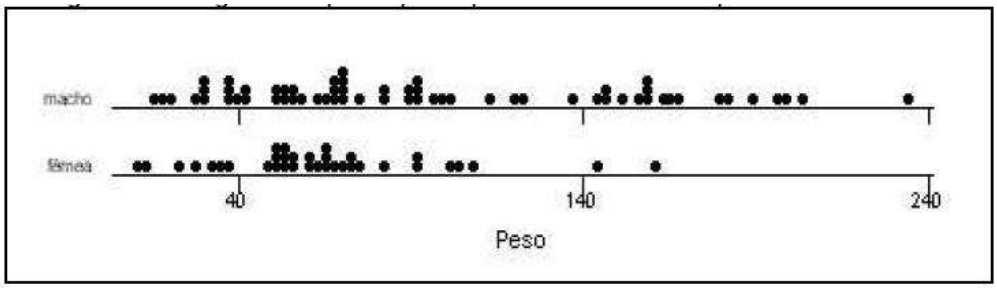


Figura 16: Diagrama de pontos para o peso de ursos machos e peso dos ursos fêmeas.

Através da comparação dos diagramas de pontos da Figura 16, podemos ver que os ursos machos possuem pesos menos homogêneos (mais dispersos) do que as fêmeas, que estão concentradas na parte esquerda do eixo de valores de peso.

O Diagrama de Ramo-e-Folhas

Outro gráfico útil e simples para representar a distribuição de freqüências de uma variável quantitativa com poucas observações é o diagrama de ramo-e-folhas. A sua sobre os demais é que ele explicita os valores dos dados, como veremos.

Exemplo dos ursos marrons (continuação):
Dos 35 ursos fêmeas observados, somente 20 puderam ter sua idade estimada. Para visualizar a distribuição dos valores de idade dessas fêmeas, usaremos um diagrama de ramo-efolhas, já que um histograma resumiria mais ainda algo que já está resumido.

Os 20 valores de idade (em meses) disponíveis, já ordenados são:

```
8 9 11 17 17 19 20 44 45 53 57 57 57 58 70 81 82 83 100 104
```

Podemos organizar os dados, separando-os pela dezenas, uma em cada linha:

```
  8   9
 11  17 17 19
 20
 44  45
 53  57 57 57 58
 70
 81  82 83
100 104
```

Como muitos valores em cada linha tem as dezenas em comum, podemos colocar as dezenas em evidência , separando-as das unidades por um traço. Ao dispor os dados dessa maneira, estamos construindo um diagrama de ramo-e-folhas (Figura 17). O lado com as dezenas é chamado de ramo, no qual estão dependuradas as unidades, chamadas folhas.

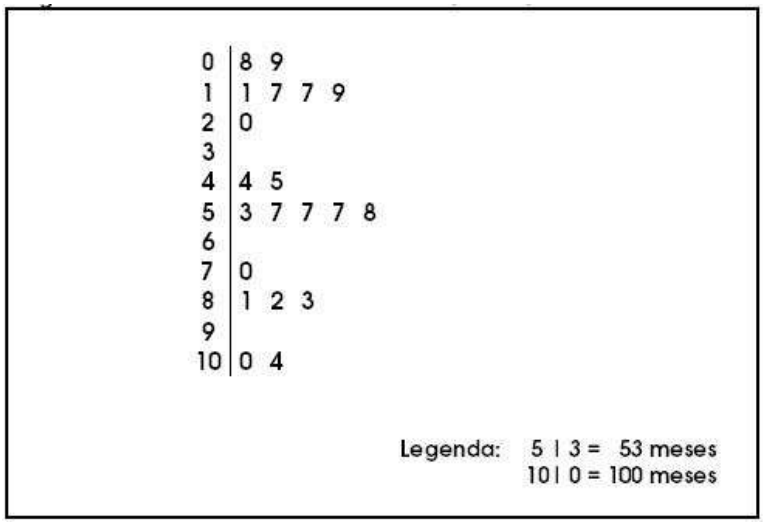


Figura 17: Ramo-e-folhas da idade (meses) dos ursos fêmeas.

Os ramos e as folhas podem representar quaisquer unidades de grandeza (dezenas e unidades, centenas e dezenas, milhares e centenas, etc). Para sabermos o que está sendo representado, um ramo-e-folhas deve ter sempre uma legenda, indicando o que significam os ramos e as folhas.

Se a idade estivesse medida em dias, por exemplo, usando esse mesmo ramo-efolhas, poderíamos estabelecer que o ramo representaria as centenas e as folhas, as dezenas. Assim, 0—8 seria igual a 80 dias e 10—4 seria igual a 1040 dias.

Analisando o ramo-e-folhas para a idade dos ursos fêmeas, percebemos a existência de três grupos: fêmeas mais jovens (até 20 meses), fêmeas mais crescidas (de 44 a 58 meses) e um grupo mais velho (mais de 70 meses), com destaque para duas fêmeas bem mais velhas.

O ramo-e-folhas também pode ser usado para comparar duas distribuições de valores, como mostra a Figura 18. Aproveitando o mesmo ramo do diagrama das fêmeas, podemos fazer o

diagrama dos machos, utilizando o lado esquerdo. Observe que as folhas dos ursos machos são dependuradas de modo espelhado, assim como explica a legenda, que agora deve ser dupla.

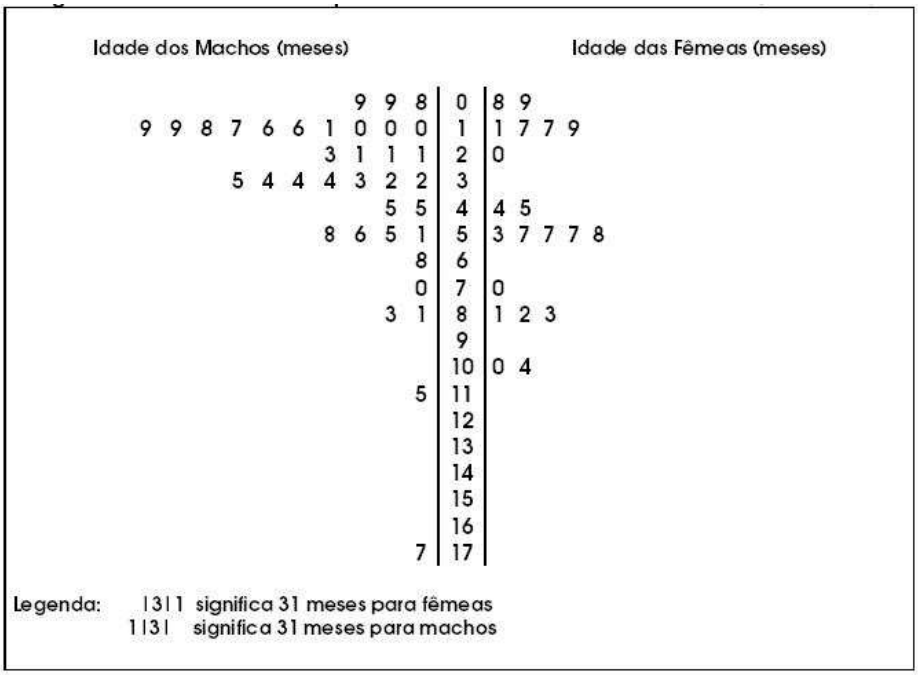


Figura 18: Ramo-e-folhas da idade (meses) dos ursos fêmeas.

Observando a Figura 18, notamos que os ursos machos são, em geral, mais jovens do que os ursos fêmeas, embora possuam dois ursos bem idosos em comparação com os demais.

Importante: No ramo-e-folhas, estamos trabalhando, implicitamente, com freqüências absolutas. Assim, ao comparar dois grupos de tamanhos diferentes, devemos levar isso em conta. Caso os tamanhos dos grupos sejam muito diferentes, não se deve adotar o ramo-e-folhas como gráfico para comparação de distribuições.

3.3.5 Aspectos Gerais da Distribuição de Freqüências

Ao estudarmos a distribuição de freqüências de uma variável quantitativa, seja em um grupo apenas ou comparando vários grupos, devemos verificar basicamente três características:

- Tendência Central;
- Variabilidade;
- Forma.

O histograma (ou o diagrama de pontos, ou o ramo-e-folhas) permite a visualização destas características da distribuição de freqüências, como veremos a seguir. Além disso, elas podem ser quantificadas através das medidas de síntese numérica (não discutidas aqui).

Tendência Central

A tendência central da distribuição de freqüências de uma variável é caracterizada pelo valor (ou faixa de valores) típico da variável.

Uma das maneiras de representar o que é típico é através do valor mais freqüente da variável, chamado de **moda**. Ou, no caso da tabela de freqüências, a classe de maior freqüência, chamada de **classe modal**. No histograma, esta classe corresponde àquela com barra mais alta ("pico").

No exemplo dos ursos marrons (Figura 13), a classe modal do peso dos ursos fêmeas é claramente a terceira, de 50 a 75 kg. Assim, os ursos fêmeas pesam, tipicamente, de 50 a 75 kg. Entretanto, para os ursos machos, temos dois picos: de 50 a 75 kg e de 150 a 175 kg. Ou seja, temos um grupo de machos com peso típico como o das fêmeas e outro grupo, menor, formado por ursos tipicamente maiores.

Dizemos que a distribuição de freqüências do peso dos ursos fêmeas é **unimodal** (apenas uma moda) e dos ursos machos é **bimodal** (duas modas). Geralmente, um histograma bimodal indica a existência de dois grupos, com valores centrados em dois pontos diferentes do eixo de valores. Uma distribuição de freqüências pode também ser amodal, ou seja, todos os valores são igualmente freqüentes.

Variabilidade

Para descrever adequadamente a distribuição de freqüências de uma variável quantitativa, além da informação do valor representativo da variável (tendência central), é necessário dizer também o quanto estes valores variam, ou seja, o quão **dispersos** eles são.

De fato, somente a informação sobre a tendência central de um conjunto de dados não consegue representá-lo adequadamente.

A Figura 19 mostra um diagrama de pontos para os tempos de espera de 21 clientes de dois bancos, um com fila única e outro com fila múltipla, com o mesmo número de atendentes. Os tempos de espera nos dois bancos têm a mesma tendência central de 7 minutos. Entretanto, os dois conjuntos de dados são claramente diferentes, pois os valores são muito mais dispersos no banco com fila múltipla.

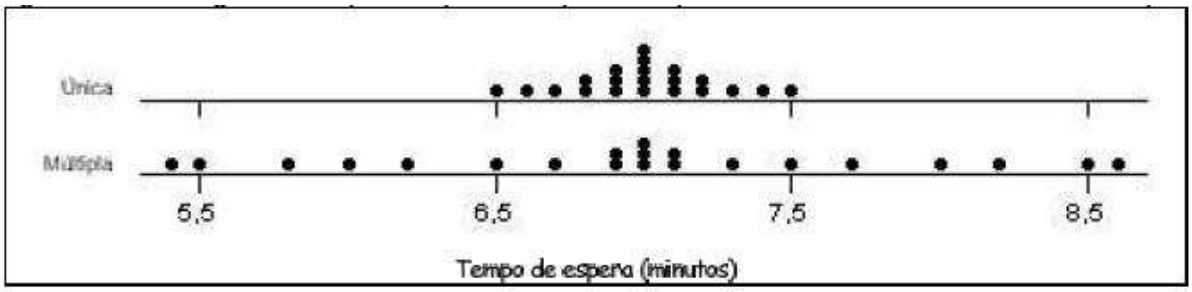


Figura 19: Ramo-e-folhas dos tempos de espera (minutos) dos clientes.

Assim, quando entramos num fila única, esperamos ser atendidos em cerca de 7 minutos, com uma **variação** de, no máximo, meio minuto a mais ou a menos. Na fila múltipla, a variação é

maior, indicando-se que tanto pode-se esperar muito mais ou muito menos que o valor típico de 7 minutos.

Forma

A distribuição de freqüências de uma variável pode ter várias formas, mas existem três formas básicas, apresentadas na Figura 20 através de histogramas e suas respectivas ogivas.

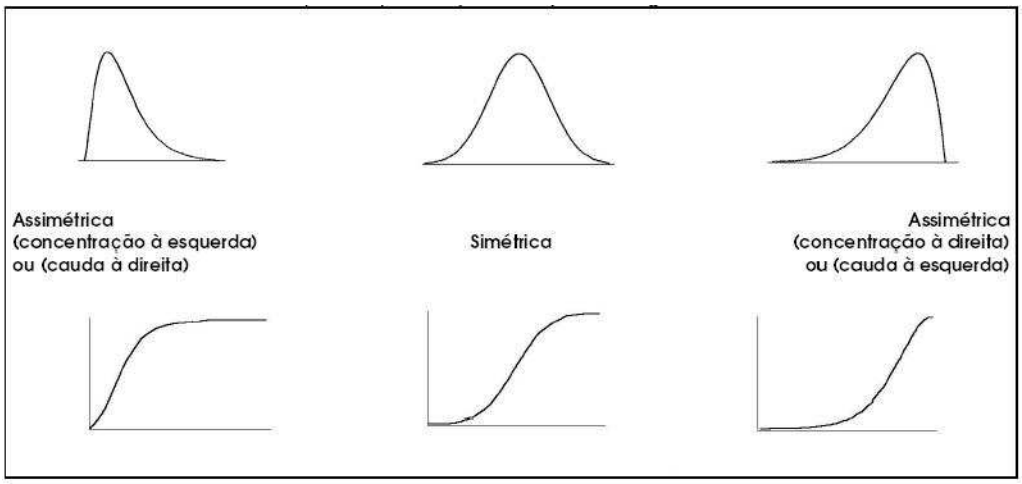


Figura 20: Ramo-e-folhas da idade (meses) dos ursos fêmeas.

Quando uma distribuição é **simétrica** em torno de um valor (o mais freqüente), significa que as observações estão igualmente distribuídas em torno desse valor (metade acima e metade abaixo).

A **assimetria** de uma distribuição pode ocorrer de duas formas:

- quando os valores concentram-se à esquerda (assimetria com concentração à esquerda ou assimetria com cauda à direita);
- quando os valores concentram-se à direita (assimetria com concentração à direita ou com assimetria cauda à esquerda);

Ao definir a assimetria de uma distribuição, algumas pessoas preferem se referir ao lado onde está a concentração dos dados. Porém, outras pessoas preferem se referir ao lado onde está faltando dados (cauda). As duas denominações são alternativas.

Em alguns casos, apenas o conhecimento da forma da distribuição de freqüências de uma variável já nos fornece uma boa informação sobre o comportamento dessa variável.

Por exemplo, o que você acharia se soubesse que a distribuição de freqüências das notas da primeira prova da disciplina de Estatística que você está cursando é, geralmente, assimétrica com concentração à direita? Como você acha que é a forma da distribuição de freqüências da renda no Brasil?

Note que, quando a distribuição é assimétrica com concentração à esquerda, a ogiva cresce bem rápido, por causa do acúmulo de valores do lado esquerdo do eixo. Por outro lado, quando a

distribuição é assimétrica com concentração à direita, o ogiva cresce lentamente no começo e bem rápido na parte direita do eixo, por causa do acúmulo de valores desse lado. Quando a distribuição é simétrica, a ogiva tem a forma de um S suave e simétrico.

A ogiva para uma distribuição de freqüências bimodal (Figura 21) mostra essa característica da distribuição através de um platô ("barriga") no meio da ogiva. A ogiva para o peso dos ursos machos (Figura 15) também mostra essa barriga .

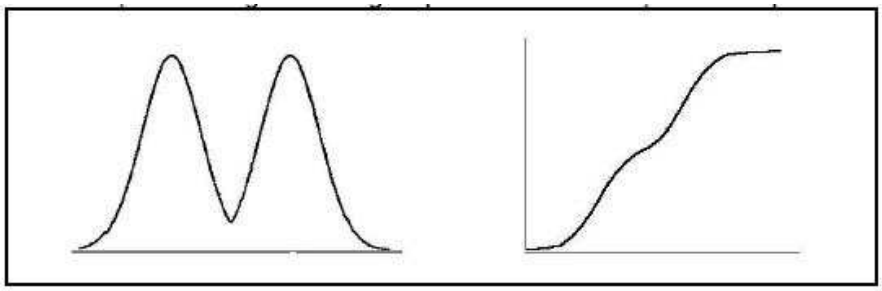


Figura 21: Ramo-e-folhas da idade (meses) dos ursos fêmeas.

Séries Temporais

Séries temporais (ou séries históricas) são um conjunto de observações de uma mesma variável quantitativa (discreta ou contínua) feitas ao longo do tempo.

O conjunto de todas as temperaturas medidas diariamente numa região é um exemplo de série temporal.

Um dos objetivos do estudo de séries temporais é conhecer o comportamento da série ao longo do tempo (aumento, estabilidade ou declínio dos valores). Em alguns estudos, esse conhecimento pode ser usado para se fazer previsões de valores futuros com base no comportamento dos valores passados.

A representação gráfica de uma série temporal é feita através do **gráfico de linha**, como exemplificado na Figura 22.

No eixo horizontal do gráfico de linha, está o indicador de tempo e, no eixo vertical, a variável a ser representada. As linhas horizontais pontilhadas são opcionais e só devem ser colocadas quando ajudarem na interpretação do gráfico. Caso contrário, devem ser descartadas, pois, como já enfatizamos antes, um gráfico deve ser o mais limpo possível.

No gráfico da Figura 22, podemos notar que a taxa de mortalidade infantil na região Nordeste esteve sempre acima da taxa da região Sudeste durante todo o período considerado, com um declínio das taxas nas duas regiões e também no Brasil como um todo ao longo do período.

Embora o declínio absoluto na taxa da região Nordeste tenha sido maior (aproximadamente 20 casos em mil nascidos vivos), a redução percentual na taxa da região Sudeste foi maior (cerca de 8 casos a menos nos 30 iniciais, ou seja, 27% a menos, enquanto 20 casos a menos nos 80 iniciais na região Nordeste representam uma redução de 25%.

Podemos observar ainda uma tendência à estabilização da taxa de mortalidade infantil da região Sudeste a partir do ano de 1994, enquanto a tendência de declínio permanece na região Nordeste

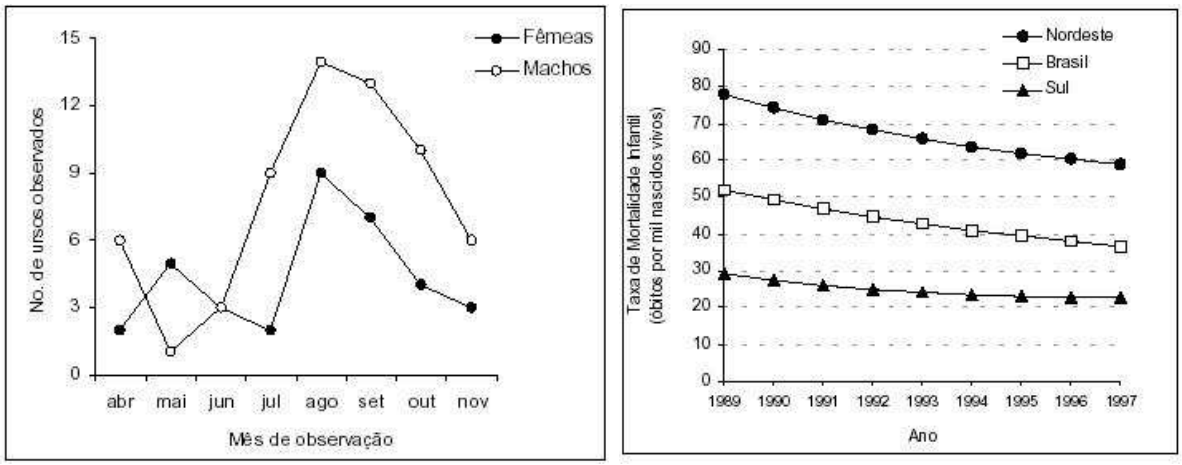


Figura 22: Gráfico de linha para o número de ursos machos e fêmeas observados ao longo dos meses de pesquisa (à esquerda) e taxa de mortalidade infantil de 1989 a 1997 nas Regiões Nordeste e Sul e no Brasil (à direita).

e no Brasil.

Ao analisar e construir um gráfico de linhas, devemos estar atentos a certos detalhes que podem mascarar o verdadeiro comportamento dos dados.

A Figura 23(a) apresenta um gráfico de linhas para o preço médio do litro de leite entre os meses de maio e agosto de 2001. Apesar de colocar os valores para cada mês, o gráfico não mostra a escala de valores e não representa a série desde o começo da escala, o valor zero.

Essa concentração da visualização da linha somente na parte do gráfico onde os dados estão situados distorce a verdadeira de dimensão da queda do preço, acentuando-a. Ao compararmos com o gráfico da Figura 23(b), cujo escala vertical começa no zero, percebemos que houve mesmo uma queda, mas não tão acentuada quanto aquela mostrada no gráfico divulgado no jornal.

Outro aspecto mascarado pela falta da escala é que as diferenças entre os valores numéricos não correspondem às distâncias representadas no gráfico.

Por exemplo, no gráfico de linha divulgado para a série do preço do leite, vemos que a queda no preço de maio para junho foi de R$0,02 e, de julho para agosto, foi de R$0,04, duas vezes maior. No entanto, a distância (vertical) entre os pontos de maio e julho é maior do que a distância (vertical) entre os pontos de julho e agosto!!

E mais, a queda de junho para junho foi de R$0,05, pouco mais do que a queda de R$0,04 de junho a agosto. Porém, a distância (vertical) no gráfico entre os pontos de junho e julho é cerca de quatro vezes maior do que a distância (vertical) dos pontos de julho e agosto!!

Examinando o gráfico apenas visualmente, sem nos atentar para os números, tenderemos a pensar que as grandes quedas no preço do leite ocorreram no começo do período de observação (de maio a julho), enquanto, na verdade, as quedas se deram quase da mesma forma mês a mês, sendo um pouco maiores no final do período (de julho a agosto).

(a) 

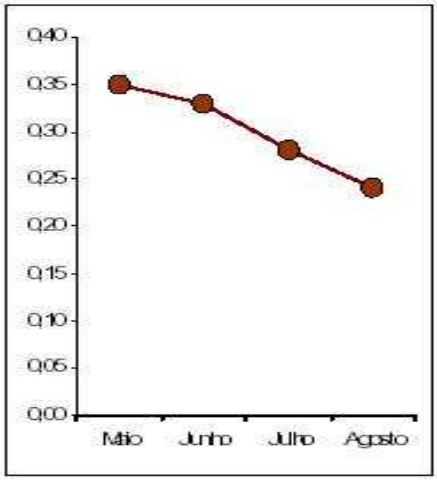
(b) 

Figura 23: Gráfico de linhas para o preço médio do litro de leite: (a) original (jornal Folha de São Paulo, set/2001), (b) modificado, com a escala de valores mostrada e iniciando-se no zero.

Além disso, a palavra despenca nos faz pensar numa queda abrupta, que é o que o gráfico divulgado parece querer mostrar. No entanto, analisando o gráfico da Figura 23(b), que corrige essas distorções, notamos que houve sim uma queda, mas não tão abrupta quanto colocada na Figura 23(a).

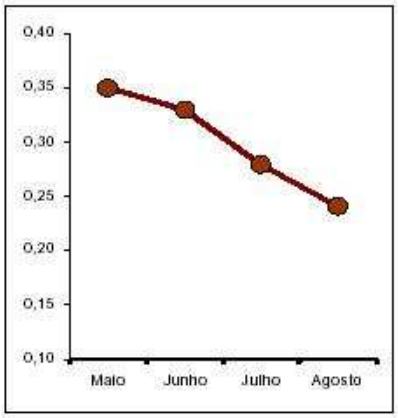
A Figura 24 mostra os efeitos na representação de uma série temporal quando mudamos o começo da escala de valores do eixo vertical. À medida que aproximamos o começo da escala do valor mínimo da série, a queda nos parece mais abrupta. A mesma observação vale para o caso em que o gráfico mostrar um aumento dos valores da série: quanto mais o início da escala se aproxima do valor mínimo da série, mais acentuado parecerá o aumento.

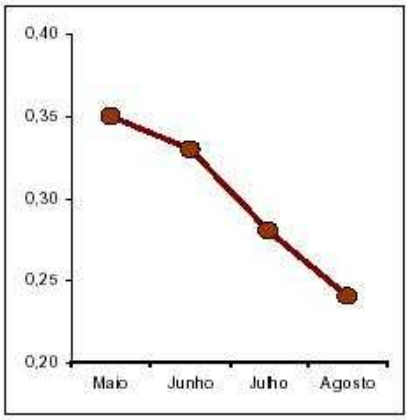
De maneira geral, um gráfico de linhas deve ser construído de modo que:

- O início do eixo vertical seja o valor mínimo possível para a variável que está sendo representada (para o caso do preço de leite, o valor zero, leite de graça), para evitar as distorções ilustradas na Figura 23;
- O final do eixo vertical seja tal que a série fica centrada em relação ao eixo vertical, como mostrado na Figura 24(a);
- Os tamanhos dos eixos sejam o mais parecidos possível, para que não ocorra a distorção mostrada nos gráficos (b) e (c)) da Figura 24.

3.3.6 O Diagrama de Dispersão

O diagrama de dispersão é um gráfico onde pontos no espaço cartesiano XY são usados para representar simultaneamente os valores de duas variáveis quantitativas medidas em cada elemento do conjunto de dados.

(a) 

(b) 

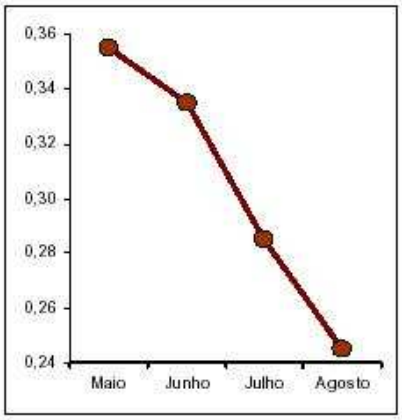
(c) 

Figura 24: Efeitos da mudança no início e/ou final da escala do gráfico em linhas da série temporal do preço do leite.

A Tabela 7 e a Figura 26 mostram um esquema do desenho do diagrama de dispersão. Neste exemplo, foram medidos os valores de duas variáveis quantitativas, X e Y, em quatro indivíduos. O eixo horizontal do gráfico representa a variável X e o eixo vertical representa a variável Y.

Tabela 7: Dados esquemáticos.

Indivíduos	Variável X	Variável Y
A	2	3
B	4	3
C	4	5
D	8	7

O diagrama de dispersão é usado principalmente para visualizar a relação/associação entre duas variáveis, mas também para é muito útil para:

- Comparar o efeito de dois tratamentos no mesmo indivíduo.
- Verificar o efeito tipo antes/depois de um tratamento;

A seguir, veremos quatro exemplos da utilização do diagrama de dispersão. Os dois primeiros referem-se ao estudo da associação entre duas variáveis. O terceiro utiliza o diagrama de dispersão para comparar o efeito de duas condições no mesmo indivíduo. O último exemplo, similar ao terceiro, verifica o efeito da aplicação de um tratamento, comparando as medidas antes e depois da medicação.

Exemplo dos ursos marrons (continuação):
Recorde que um dos objetivos dos pesquisadores neste estudo é encontrar uma maneira de conhecer o peso do urso através de uma medida mais fácil de se obter do que a direta (carregar

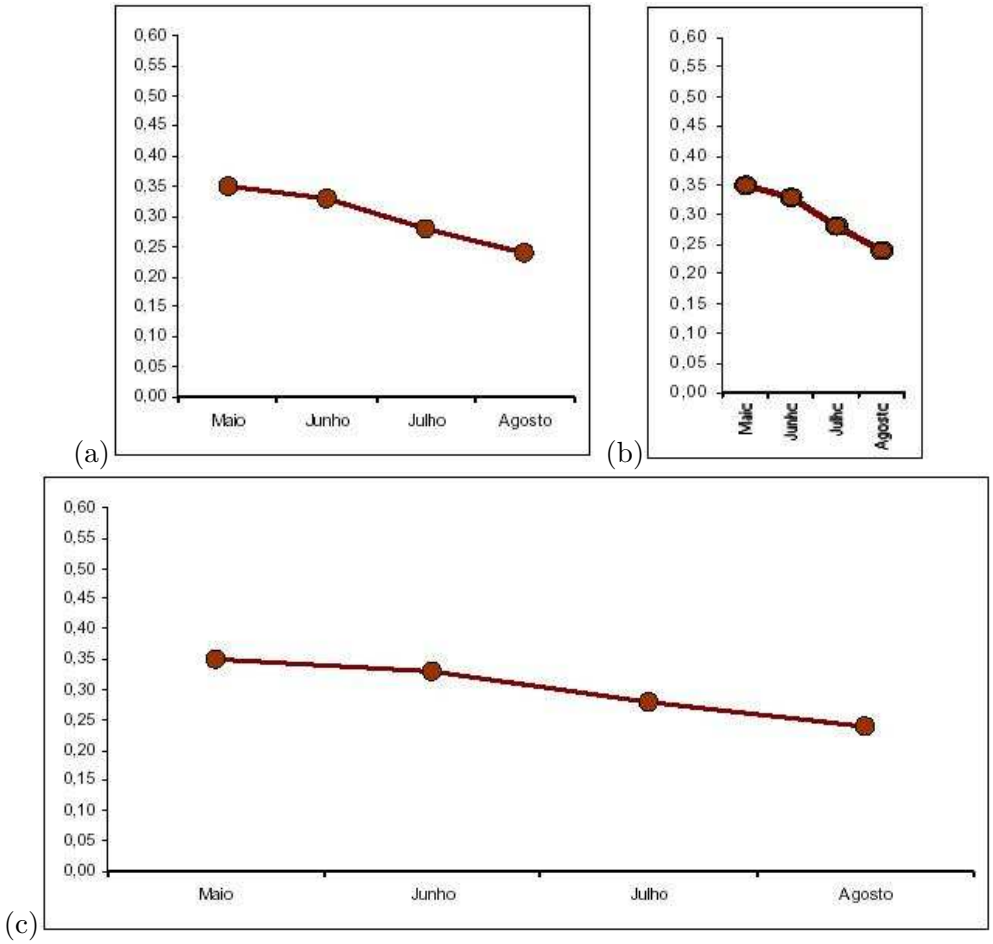


Figura 25: Efeitos de alterações na dimensão horizontal do gráfico de linhas da série do preço do leite.

uma balança para o meio da selva e colocar os ursos em cima dela) como, por exemplo, uma medida de comprimento (altura, perímetro do tórax, etc.).

O problema estatístico aqui é encontrar uma variável que tenha uma relação forte com o peso, de modo que, a partir de seu valor medido, possa ser calculado (estimado, na verdade) o valor peso indiretamente, através de uma equação matemática.

O primeiro passo para encontrar esta variável é fazer o diagrama de dispersão das variáveis candidatas (eixo horizontal) versus o peso (eixo vertical), usando os pares de informações de todos os ursos. Você pode tentar as variáveis: idade, altura, comprimento da cabeça, largura da cabeça, perímetro do pescoço e perímetro do tórax.

Na Figura 27, mostramos a relação entre peso e altura e entre peso e perímetro do tórax. Respectivamente.

Podemos ver que, tanto a altura quanto o perímetro do tórax são fortemente associados ao peso do urso, no sentido de que quanto mais alto o urso ou quanto maior a medida de seu tórax, mais pesado ele será.

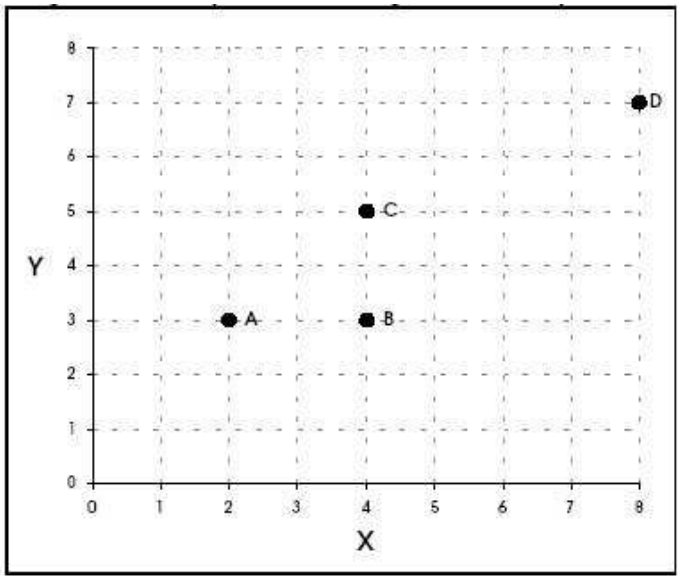


Figura 26: Esquema do diagrama de dispersão.

Mas note que este crescimento é linear para o perímetro do tórax e não-linear para a altura.

Além disso, com os pontos estão mais dispersos no gráfico da altura, a variável mais adequada para estimar, sozinha, o peso é o perímetro do tórax (a técnica estatística adequada aqui chama-se Regressão Linear Simples).

Exemplo dos morangos:
Um produtor de morangos para exportação deseja produzir frutos grandes, pois frutos pequenos têm pouco valor mesmo no mercado interno. Além disso, os frutos, mesmo grandes, não devem ter tamanhos muito diferentes entre si. O produtor suspeita que uma dos fatores que altera o tamanho dos frutos é o número de frutos por muda.

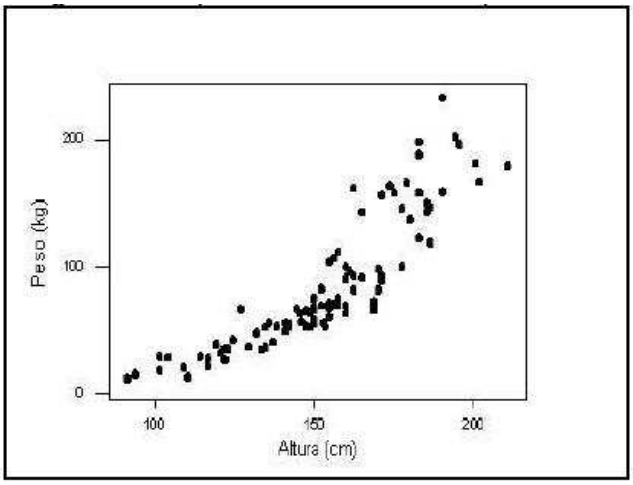
Para investigar a relação entre o número de frutos que uma planta produz e o peso destes frutos, ele observou dados de 10 morangueiros na primeira safra (Tabela 8). O diagrama de dispersão é mostrado na Figura 28.

O diagrama de dispersão mostra-nos dois fatos. O primeiro, que há um decréscimo no valor médio do peso do fruto por árvore à medida que cresce o número de frutos na árvore. Ou seja, não é vantagem uma árvore produzir muitos frutos, pois ele tenderão a ser muito pequenos.

O segundo fato que percebemos é que, com o aumento no número de frutos na árvores, cresce também a variabilidade no peso, gerando tanto frutos muito grandes, como muito pequenos.

Assim, conclui-se que não é vantagem ter poucas plantas produzindo muito frutos, mas sim muitas plantas produzindo poucos frutos, mas grandes e uniformes. Uma análise mais detalhada poderá determinar o número ideal de frutos por árvore, aquele que maximiza o peso médio e, ao mesmo tempo, minimiza a variabilidade do peso.

Exemplo da Capacidade Pulmonar:

(a) 

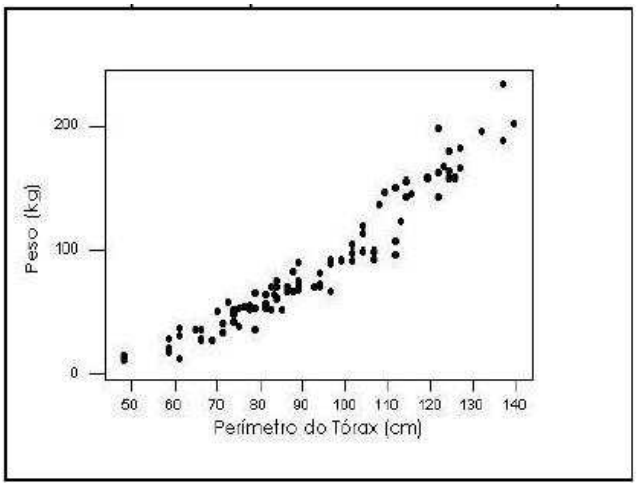
(b) 

Figura 27: Diagrama de dispersão da altura versus o peso (a) e do perímetro do tórax versus o peso (b) dos ursos marrons.

Tabela 8: Peso dos frutos e número de frutos por planta em 10 morangueiros na primeira safra.

Muda	N	Peso dos Frutos (gramas)													
1	5	15,2	15,5	15,6	15,7	16,4									
2	6	14,0	14,5	15,4	15,9	15,9	16,1								
3	7	13,7	13,8	14,1	14,1	14,5	14,9	15,5							
4	8	11,0	11,5	12,4	12,4	12,9	14,5	15,5	16,6						
5	9	10,2	11,1	12,1	12,4	13,5	13,8	14,0	15,4	16,0					
6	10	9,0	9,3	10,7	11,6	11,7	12,6	12,8	12,8	13,4	15,1				
7	11	7,8	8,6	8,7	9,6	11,1	11,9	12,1	12,5	14,1	14,2	14,0			
8	12	7,3	9,4	10,2	10,3	10,8	10,6	11,1	11,5	11,5	12,9	13,4	15,0		
9	13	6,9	7,6	8,5	10,0	10,9	11,0	11,4	11,6	12,0	12,0	12,7	13,5	14,0	
10	14	7,0	8,0	9,0	10,0	10,0	10,5	11,0	11,2	11,2	11,7	12,5	12,9	13,5	13,5

Captopril é um remédio destinado a baixar a pressão sistólica. Para testar seu efeito, ele foi ministrado a 12 pacientes, tendo sido medida a pressão sistólica antes e depois da medicação (Tabela 9).

Os mesmos indivíduos foram utilizados nas duas amostras (Antes/depois). Assim, é natural compararmos a pressão sistólica para cada indivíduo, comparando a pressão sistólica depois e antes. Para todos os pacientes, a pressão sistólica depois do Captopril é menor do que antes da medicação. Mas como podemos ver se estas diferenças são grandes ? Através do diagrama de dispersão mostrado na Figura 29.

Cada ponto no diagrama de dispersão corresponde às medidas de pressão sistólica de um paciente, medida antes e depois da medicação.

A linha marcada no diagrama corresponde à situação onde a pressão sistólica não se alterou

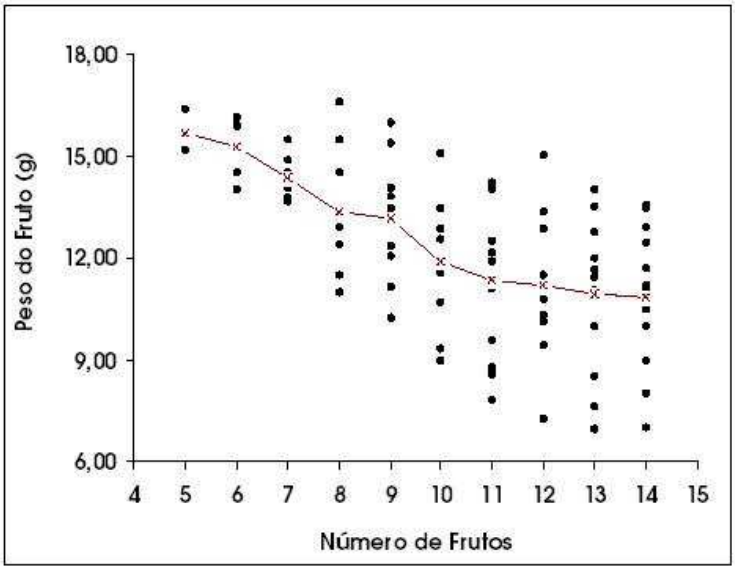


Figura 28: Diagrama de dispersão do número de frutos por árvore versus o peso do fruto e linha unindo os pesos médios dos frutos.

Tabela 9: Pressão sistólica (mmHg) medida em 12 pacientes antes e depois do Captopril.

Paciente	A	B	C	D	E	F	G	H	I	J	K	L
Antes	200	174	198	170	179	182	193	209	185	155	169	210
Depois	191	170	177	167	159	151	176	183	159	145	146	177

depois do paciente tomar o Captopril.

Veja que todos os pontos estão abaixo desta linha, ou seja para todos os pacientes o Captopril fez efeito. Grande parte destes pontos está bem distante da linha, mostrando que a redução na pressão sistólica depois do uso do medicamento não foi pequena.

3.3.7 O Ladder Plot

O **ladder plot** não é um gráfico do tipo padrão mas pode ser útil para visualizar **dados pareados**. Considere o seguinte exemplo:

Um ornitologista deseja saber se um determinado local é usado por pássaros migratórios de uma certa raça para engorda antes de migrar.

Ele captura alguns pássaros em Agosto e pesa-os, então em Setembro ele tenta re-capturar os mesmos pássaros e faz novas medidas. Ele re-capturou 10 dos pássaros duas vezes, ambos em Agosto e Setembro.

A tabela abaixo mostra as massas desses pássaros.

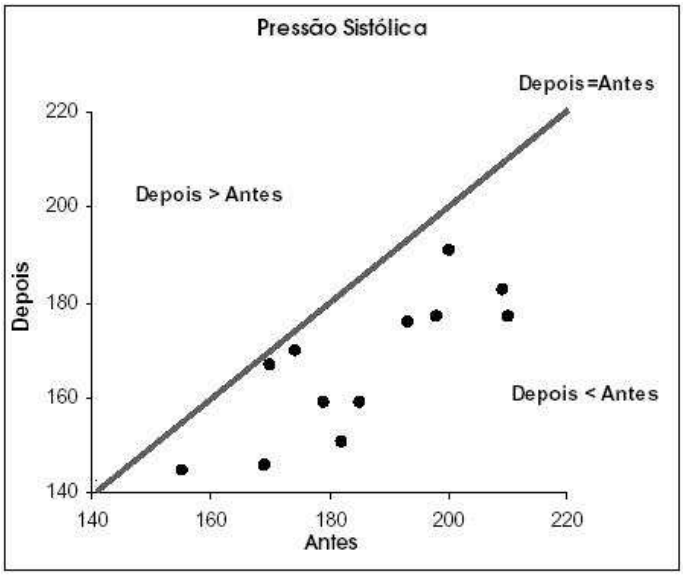


Figura 29: Diagrama de dispersão da pressão sistólica antes X depois da medicação e linha correspondendo ao não efeito individual da medicação.

Mass in August (g)	Mass in September (g)
10.3	12.2
11.4	12.1
10.9	13.1
12.0	11.9
10.0	12.0
11.9	12.9
12.2	11.4
12.3	12.1
11.7	13.5
12.0	12.3

O ladder plot destes dados fica como segue (Figura 30):

É muito mais fácil ver do gráfico do que da tabela que os pássaros tendem a engordar, e que aqueles que não engordaram tenderam a ser os maiores que provavelmente não necessitam de uma engorda extra.

3.3.8 Dados múltiplos

Os resultados de um estudo tipicamente envolverão mais do que uma única amostra de dados.

Representações gráficas são úteis para comparar grupos de dados ou para verificar se exitem relações entre eles.

Existem muitas possibilidades, mas a mais adequada dependerá das peculiaridades de cada conjunto de dados.

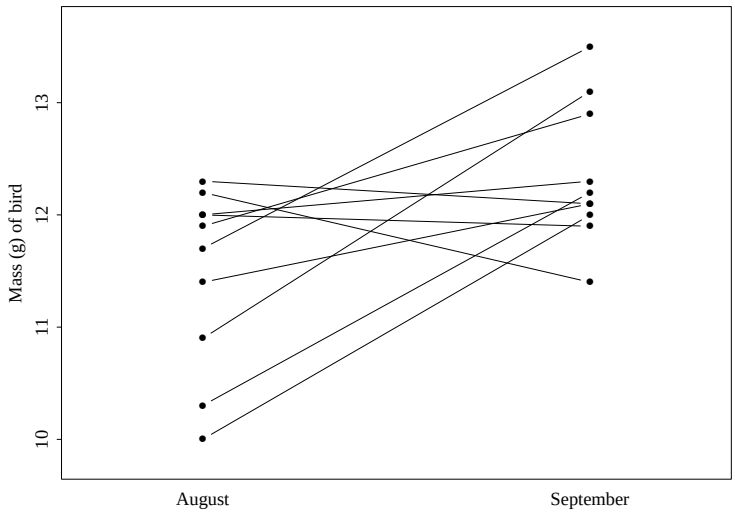


Figura 30: Pesos de pássaros em duas ocasões

Além dos exemplos abaixo, podemos criar combinações de métodos já discutidos. Por exemplo, se medirmos as alturas e pesos de uma amostra de pessoas, podemos produzir:

- box-plots de altura lado a lado para homens e mulheres,
- gráficos ramo-e-folhas lado a lado (com as alturas dos homens à esquerda do ramo, e as alturas das mulheres à direita),
- um histograma acima do outro (com a mesma escala no eixo x de forma que eles possam ser facilmente comparados).

Para um número diferente de grupos, uma série de box-plots verticais funciona bem como um símples resumo dos dados.

Para combinações de dados categóricos, uma série de gráficos de setores podem ser produzidos, i.e. dois gráficos de setores, um para homens e um para mulheres.

4 Estatística Descritiva - Medidas Resumo

4.1 Dados qualitativos

Para sumarizar dados qualitativos numericamente, utiliza-se **contagens, proporções, percentagens, taxas por 1000, taxas por 1.000.000**, etc, dependendo da escala apropriada.

Por exemplo, se encontrarmos que 70 de 140 estudantes de medicina são homens, poderíamos relatar o resultado como uma proporção (0.5) ou provavelmente como um percentual (50%) ou ainda como uma taxa (1 a cada 2 estudantes da medicina são do sexo masculino).

Se encontrarmos que 7 de uma amostra de 5000 pessoas são portadores de uma doença rara poderíamos expressar isto como uma proporção observada (0.0014) ou percentual (0.14%), mas melhor seria 1.4 casos por mil.

4.1.1 Resumindo numericamente

Considere a seguinte tabela que mostra a distribuição de 100 pacientes quanto à tipagem sanguínea.

A	8
B	33
AB	32
O	17
Total	100

A moda de um conjunto de dados categóricos é a categoria que ocorre com maior frequencia. Ela deve ser usada cuidadosamente como uma medida resumo global porque é muito dependente da forma como os dados são categorizados. Para os dados dos sexos dos ursos marrons a moda é *machos*. Para os dados acima, a categoria modal é "sangue tipo B", mas por muito pouco.

4.2 Dados quantitativos

4.2.1 Resumindo numericamente

Para resumir numericamente dados quantitativos o objetivo é escolher medidas apropriadas de **locação** ("qual o tamanho dos números envolvidos?") e de **dispersão** ("quanta variação existe?") para os tipos de dados.

Existem três escolhas principais para a medida de locação, a chamada "3 Ms", as quais estão ligadas a certas medidas de dispersão como segue:

M	'Dispersão'
média (o valor 'médio')	desvio padrão
mediana (o valor do 'meio')	IQR
moda (o valor 'mais comum')	proporção

4.2.2 A moda

Nem todos os conjuntos de dados são suficientemente balanceados para o cálculo da média ou mediana. Algumas vezes, especialmente para dados de contagem, um único valor domina a amostra.

A medida de locação apropriada é então a **moda**, a qual é o valor que ocorre com maior frequência. A proporção da amostra a qual toma este valor modal deveria ser utilizada no lugar de uma medida formal de dispersão.

Algumas vezes, podemos distinguir claramente dois ou mais 'picos' na frequência dos valores registrados. Neste caso (chamado **bimodal/multimodal**) deveríamos apresentar ambas as localizações. Dados deste tipo são particularmente difíceis de resumir (e analisar).

Exemplo. Dez pessoas registraram o número de copos de cerveja que eles tomaram num determinado sábado:

$$0, 0, 0, 0, 0, 1, 2, 3, 3, 6$$

A moda é 0 copos de cerveja, a qual foi obtida pela metade da amostra. Poderiamos adicionar mais informação separando a amostra e dizendo que daqueles que tomaram cerveja a mediana foi de 3 copos.

4.2.3 A mediana e a amplitude inter-quartis

Uma outra forma de sumarizar dados é em termos dos **quantis** ou **percentis**. Essas medidas são particularmente úteis para dados não simétricos.

A **mediana** (ou percentil 50) é definida como o valor que divide os dados ordenados ao meio, i.e. metade dos dados têm valores maiores do que a mediana, a outra metade tem valores menores do que a mediana.

Adicionalmente, os quartis **inferior** e **superior**, Q1 e Q3, são definidos como os valores abaixo dos quais estão um quarto e três quartos, respectivamente, dos dados.

Estes três valores são frequentemente usados para resumir os dados juntamente com o mínimo e o máximo.

Eles são obtidos ordenando os dados do menor para o maior, e então conta-se o número apropriado de observações: ou seja é $\frac{n+1}{4}$, $\frac{n+1}{2}$ e $\frac{3(n+1)}{4}$ para o quartil inferior, mediana e quartil superior, respectivamente.

Para um número par de observações, a mediana é a média dos valores do meio (e analogamente para os quartis inferior e superior).

A medidade de dispersão é a **amplitude inter-quartis**, IQR = Q3 − Q1, i.e. é a diferença entre o quartil superior e o inferior.

Exemplo. O número de crianças em 19 famílias foi

0, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 4, 4, 5, 6, 6, 7, 8, 10

A **mediana** é o (19+1) / 2 = 10^o valor, i.e. 3 crianças.

O quartil **inferior** e **superior** são os valores 5^o e 15^o, i.e. 2 e 6 crianças, portanto **amplitude inter-quartil** é de 4 crianças. Note que 50% dos dados estão entre os quartis inferior e superior.

Box-and-Whisker Plots

Box-and-Whisker plots ou simplesmente **box-plots** são simples representações diagramáticas dos cinco números sumários: (mínimo, quartil inferior, mediana, quartil superior, máximo).

Um box-plot para os dados acima fica como mostrado a seguir (Figura 31).

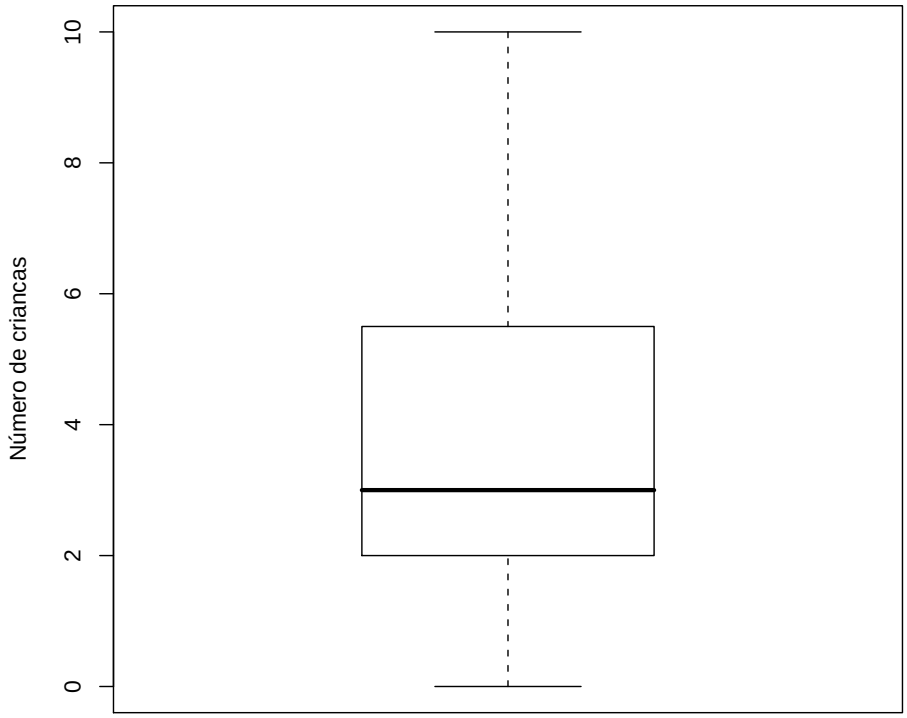


Figura 31: Representação dos 5 números sumários num box-plot

4.2.4 Média, variância e desvio padrão

Para resumir dados quantitativos aproximadamente **simétricos**, é usual calcular a **média** aritmética como uma medida de locação. Se $x_1, x_2, \ldots, x_n$ são os valores dos dados, então podemos escrever a média como

$$\overline{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n},$$

onde '$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$' e frequentemente é simplificada para $\sum x_i$ ou até mesmo $\sum x$ que significa 'adicione todos os valores de x'.

A **variância** é definida como o 'desvio quadrático médio da média' e é calculada de uma amostra de dados como

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

A segunda versão é mais fácil de ser calculada, no entanto muitas calculadoras têm funções prontas para o cálculo de variâncias, e é raro ter que realisar todos os passos manualmente.

Comumente as calculadoras fornecerão a raiz quadrada da variância, o **desvio padrão**, i.e.

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

a qual é medida nas mesmas unidades dos dados originais.

Uma informação útil é que para qualquer conjunto de dados, pelo menos 75% deles fica dentro de uma distância de 2 desvios padrão da média, i.e. entre $\bar{x} - 2s$ e $\bar{x} + 2s$.

Exemplo. Sete homens foram pesados, e os resultados em kg foram:

$$57.0,\ 62.9,\ 63.5,\ 64.1,\ 66.1,\ 67.1,\ 73.6.$$

A **média** é $454.3/7 = 64.9$ kg,

a **variância** é $(29635.05 - 454.3^2/7)/6 = 25.16$ kg^2

e o **desvio padrão** é $\sqrt{25.16} = 5.02$ kg.

4.2.5 Coeficiente de variação

Uma pergunta que pode surgir é: *O desvio padrão calculado é grande ou pequeno?*

Esta questão é relevante por exemplo, na avaliação da precisão de métodos.

Um desvio padrão pode ser considerado grande ou pequeno dependendo da ordem de grandeza da variável.

Uma maneira de se expressar a variabilidade dos dados tirando a influência da ordem de grandeza da variável é através do **coeficiente de variação**, definido por:

$$CV = \frac{s}{\bar{x}}$$

O CV é:

- interpretado como a ***variabilidade dos dados em relação à média***. Quanto menor o CV mais homogêneo é o conjunto de dados.
- ***adimensional***, isto é, um número puro, que será positivo se a média for positiva; será zero quando não houver variabilidade entre os dados, ou seja, $s = 0$.
- usualmente ***expresso em porcentagem***, indicando o percentual que o desvio padrão é menor ($100\%CV < 100\%$) ou maior ($100\%CV > 100\%$) do que a média

Um CV é considerado baixo (indicando um conjunto de dados razoavelmente homogêneo) quando for **menor ou igual a 25%**. Entretanto, esse padrão varia de acordo com a aplicação.

Por exemplo, em medidas vitais (batimento cardíaco, temperatura corporal, etc) espera-se um CV muito menor do que 25% para que os dados sejam considerados homogêneos.

Pode ser difícil classificar um coeficiente de variação como baixo, médio, alto ou muito alto, mas este pode ser bastante útil na comparação de duas variáveis ou dois grupos que a princípio não são comparáveis.

Exemplos:

1. Em um grupo de pacientes foram tomadas as pulsações (batidas por minuto) e dosadas as taxas de ácido úrico (mg/100ml). As médias e os desvios padrão foram:

Variável	$\bar{x}$	s
pulsação	68,7	8,7
ácido úrico	5,46	1,03

   Os coeficientes de variação são: $CV_p = 8,7/68,7 = 0,127$ e $CV_{a.u.} = 1,03/5,46 = 0,232$, o que evidencia que a pulsação é mais estável do que o ácido úrico.

2. Em experimentos para a determinação de clorofila em plantas, levantou-se a questão de que se o método utilizado poderia fornecer resultados mais consistentes. Três métodos foram colocados à prova e 12 folhas de abacaxi foram analisadas com cada um dos métodos. Os resultados foram os seguintes:

Método (unidade)	$\bar{x}$	s	CV
$1(100cm^3)$	13,71	1,20	0,088
$2(100g)$	61,40	5,52	0,090
$3(100g)$	337,00	31,20	0,093

   Note que as médias são bastante diferentes devido às diferenças entre os métodos. Entretanto, os três CV são próximos, o que indica que a consistência dos métodos é praticamente equivalente, sendo que o método 3 mostrou-se um pouco menos consistente.

4.2.6 Escore padronizado

O escore padronizado, ao contrário do CV, é útil para comparação dos resultados indivíduais.

Por exemplo, um aluno que tenha obtido nota 7 numa prova cuja média da classe foi 5 foi melhor do que numa prova em que tirou 8 mas a média da classe foi 9.

Além da comparação da nota individual com a média da classe, também é importante avaliar am cada caso se a variabilidade das notas foi grande ou não.

Por exemplo, o desempenho deste aluno que obteve nota 7 seria bastante bom se o desvio padrão da classe fosse 2 e apenas razoável se o desvio padrão da classe fosse 4.

Sejam $x_1, x_2, \cdots, x_n$ os dados observados em uma amostra de tamanho n e $\bar{x}$ e s a média e o desvio padrão, então

$$z_i = \frac{x_i - \bar{x}}{s}, i = 1, \cdots, n$$

é denominado **escore padronizado**.

Os escores padronizados são muito úteis na comparação da posição relativa da medida de um indivíduo dentro do grupo ao qual pertence, o que justifica sua grande apliacação como medida de avaliação de desempenho.

Exemplo:
Os escores padronizados são amplamente utilizados em teste de aptidão física. Mathews (1980) compara testes de aptidão física e conhecimento desportivo.

Tabela 10: Resultados obtidos por duas alunas do curso secundário, média e desvio padrão da turma em teste de aptidão física e conhecimento desportivo

Teste	$\bar{x}$	s	x		z	
			Maria	Joana	Maria	Joana
abdominais em 2 min	30	6	42	38	2,00	1,33
salto em extensão (cm)	155	23	102	173	-2.33	0,78
suspensão braços flexionados (seg)	50	8	38	71	-1.50	2,63
correr/andar em 12 min (m)	1829	274	2149	1554	1,17	-1,00
conhecimento desportivo	75	12	97	70	1,83	-0,42

Maria apresentou um desempenho muito acima da média em força abdominal (dois desvio padrão acima da média); sua capacidade aeróbica (corrida/caminhada) está acima da média mas não é notável e ela tem um conhecimento desportivo bastante bom comparado com o grupo.

No salto de extensão e na suspensão com flexão do braço sobre antebraço, Maria obteve escores abaixo das respectivas médias do grupo, sendo que o desempenho de Maria para salto em extensão é bastante ruim.

Descreva o desempenho de Joana.

5 Introdução à probabilidade e aplicação em testes diagnósticos

Nesta seção serão introduzidos conceitos probabilísticos aplicados a um problema de verificação da qualidade de um teste diagnóstico.

5.1 Probabilidade

De maneira informal, probabilidade é uma medida da certeza de ocorrência de um evento. Formalmente, existem duas definições de probabilidade: a definição clássica e a frequentista.

5.1.1 Definição clássica

Considere o seguinte experimento aleatório: lançar uma moeda e observar a face voltada para cima.

Este experimento possui dois resultados possíveis: cara e coroa. Ao conjunto dos resultados possíveis de um experimento chamamos de **espaço amostral** e será denotado pela letra E. O espaço amostral do experimento acima é $E = \{c, \bar{c}\}$, em que c denota cara e $\bar{c}$ coroa.

Um subconjunto do espaço amostral é chamado de **evento** e é denotado por letras maiúsculas. Para o exemplo acima, podemos definir os eventos:
$A = \{c\} = \{\text{ocorrer cara}\}$ e $B = \{\bar{c}\} = \{\text{ocorrer coroa}\}$

O evento A acima é chamado de **evento simples** pois é constituído de apenas um elemento do espaço amostral. O mesmo se aplica para o evento B.

Seja A um evento qualquer do espaço amostral. Se os eventos simples são equiprováveis podemos calcular $P(A)$ como:

$$P(A) = \frac{\text{número de resultados favoráveis à ocorrência do evento A}}{\text{número de resultados possíveis}} \tag{1}$$

Para o experimento acima se a moeda for não viciada, os eventos A e B serão equiprováveis e $P(A) = P(B) = 1/2$.

No lançamento de um dado não viciado, os eventos simples são equiprováveis com probabilidade 1/6, $P(\text{sair um número par}) = 3/6 = 1/2$, $P(\text{sair número 1 ou 3}) = 2/6 = 1/3$ e $P(\text{sair número maior do que 2}) = 4/6 = 2/3$.

5.1.2 Definição frequentista

Na maioria das situações práticas, os eventos simples do espaço amostral não são equiprováveis e não podemos calcular probabilidades usando a definição clássica. Neste caso, vamos calcular probilidades como a frequência relativa de um evento. Segue um exemplo que ilustra o método.

Exemplo 1: Uma amostra de 6800 pessoas de uma determinada população foi classificada quanto à cor dos olhos e à cor dos cabelos. Os resultados foram:

Tabela 11: Classificação de uma amostra de 6800 pessoas quanto à cor dos olhos e à cor dos cabelos

	Cor dos cabelos				
Cor dos olhos	Loiro	Castanho	Preto	Ruivo	Total
Azul	1768	807	189	47	2811
Verde	946	1387	746	53	3132
Castanho	115	438	288	16	857
Total	2829	2632	1223	116	6800

Considere o experimento aleatório que consiste em classificar um indivíduo quanto à cor dos olhos. O espaço amostral é $E = \{A, V, C\}$, em que:

A={a pessoa tem olhos azuis}
V={a pessoa tem olhos verdes}
C={a pessoa tem olhos castanhos}

Os eventos acima não são equiprováveis. Então vamos calcular a probabilidade de ocorrer um evento como a frequência relativa deste evento:

$$P(A) = \frac{\text{número de pessoas de olhos azuis}}{\text{número de pessoas na amostra}} = \frac{2811}{6800} = 0,4134 \tag{2}$$

O valor obtido é na verdade uma **estimativa** da probabilidade. A qualidade desta estimativa depende do número de replicações do experimento, ou seja, do tamanho da amostra.

À medida que o tamanho da amostra cresce, a estimativa aproxima-se mais do valor verdadeiro da probabilidade. Vamos, no entanto, assumir que o número de replicações é suficientemente grande para que a diferença entre a estimativa e o valor verdadeiro da probabilidade seja desprezível.

As probabilidades dos eventos V e C são:

$P(V) = \frac{3132}{6800} = 0,4606$ e $P(C) = \frac{857}{6800} = 0,1260$

Observe que $P(A) + P(V) + P(C) = 1$. Este resultado é geral, uma vez que a união destes eventos corresponde ao espaço amostral.

Seja $\bar{A}$ o evento {a pessoa não tem olhos azuis}. O evento $\bar{A}$ é chamado de evento complementar de A e $P(\bar{A}) = \frac{3132+857}{6800} = 0,5866 = 1 - P(A)$.

Estes resultados são propriedades de probabilidades. Seja A um evento qualquer no espaço amostral E. Então valem as propriedades:

1. $0 \leq P(A) \leq 1$
2. $P(E) = 1$
3. $P(\bar{A}) = 1 - P(A)$

Voltando ao exemplo, vamos calcular algumas probabilidades. Seja L o evento {a pessoa tem cabelos loiros}.

Qual a probabilidade de uma pessoa ter olhos azuis e cabelos loiros?

O evento {a pessoa tem olhos azuis e cabelos loiros} é chamado de **evento interseção**. Ele contém todos os elementos do espaço amostral pertencentes concomitantemente ao evento A e ao evento L e será denotado por $A \cap L$, e a probabilidade deste evento é:

$$P(A \cap L) = \frac{1768}{6800} = 0,26 \tag{3}$$

Qual a probabilidade de uma pessoa ter olhos azuis ou cabelos louros?

O evento {a pessoa tem olhos azuis ou cabelos louros} é chamado de evento união e será denotado por $A \cup L$. Ele contém todos os elementos do espaço amostral que estão em A, ou somente em L, ou em ambos, e a probabilidade deste evento é:

$$P(A \cup L) = P(A) + P(L) - P(A \cap L) = \frac{2811}{6800} + \frac{2829}{6800} - \frac{1768}{6800} = \frac{3872}{6800} = 0,5694 \tag{4}$$

Para quaisquer dois eventos A e B do espaço amostral, podemos calcular a probabilidade do evento união da seuignte forma: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Se os eventos são **mutuamente exclusivos**, isto é, eles não podem ocorrer simultaneamente, $P(A \cap B) = 0$ e consequentemente

$$P(A \cup B) = P(A) + P(B)$$

Num exemplo de lançamento de um dado como os eventos P = {sair número par} e I = {sair número ímpar} são mutuamente exclusivos, $P(P \cup I) = P(P) + P(I) = 3/6 + 3/6 = 1$.

Entretanto, os eventos O = {sair número 1 ou 3} e Q = {sair número maior que 2} não são mutuamente exclusivos, pois $O \cap Q = \{3\}$.

Neste caso, $P(O \cup Q) = P(O) + P(Q) - P(O \cap Q) = 2/6 + 4/6 - 1/6 = 5/6$.

5.1.3 Probabilidade condicional

A probabilidade de um evento A ocorrer, dado que um outro evento B ocorreu, é chamada probabilidade condicional do evento A dado B.

Por exemplo, a probabilidade de que uma pessoa venha a contrair AIDS dado que ele/ela é um usuário de drogas injetáveis é uma probabilidade condicional.

Um outro exemplo, é um estudo sobre panfletos de supermercado, em que deseja-se calcular a probabilidade de que um panfleto de propaganda seja jogado no lixo dado que contém uma mensagem sobre o cuidado de depositar lixo no lixo.

Um terceiro exemplo, é uma frase que ocorrerá repetidamente neste material: "Se a hipótese nula for verdadeira, a probabilidade de se obter um resultado como este é ...". Aqui a palavra *se* substitui a palavra *dado que*, mas o sentido é o mesmo.

Com dois eventos, A e B, a probabilidade condicional de A dado B é denotada por $P(A|B)$, por exemplo, P(AIDS—usuário de drogas) ou P(lixo—mensagem).

Exemplo 2: Frequentemente assumimos, com alguma justificativa, que a paternidade leva a responsabilidade. Pessoas que passam anos atuando de maneira descuidadosa e irracional de alguma forma parecem se tornar em pessoas diferentes uma vez que elas se tornam pais, mudando muitos dos seus antigos padrões habituais. Suponha que uma estação de rádio tenha amostrado 100 pessoas, 20 das quais tinham crianças. Eles observaram que 30 dessas pessoas usavam cinto de segurança, e que 15 daquelas pessoas tinham crianças. Os resultaddos são mostrados na Tabela 12.

Paternidade	**Usam cinto**	**Não usam cinto**	**Total**
Com crianças	15	5	20
Sem crianças	15	65	80
Total	30	70	100

Tabela 12: Relação entre paternidade e uso de cinto de segurança.

A partir da informação na Tabela 12 podemos calcular probabilidades simples (ou marginais ou incondicionais), conjuntas e condicionais.

- A probabilidade de uma pessoa amostrada aleatoriamente usar cinto de segurança é 30/100=0,30.
- A probabilidade de uma pessoa ter criança **e** usar cinto de segurança é 15/100=0,15.
- A probabilidade de uma pessoa usar cinto de segurança **dado que** tem criança é 15/20=0,75.
- A probabilidade de uma pessoa ter criança **dado que** usa cinto de segurança é 15/30=0,50.

A probabildade condicional também pode ser obtida por:

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

Esta expressão pode ser reescrita como:

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

A probabilidade do evento $\bar{A}$ (complementar de A) dado que o evento B ocorreu, isto é, $P(\bar{A}|B)$, é expressa por:

$$P(\bar{A}|B) = 1 - P(A|B)$$

Os eventos A e B são independentes se o fato de um deles ter ocorrido não altera a probabilidade da ocorrência do outro, isto é,

$$P(A|B) = P(A) \text{ou} P(B|A) = P(B)$$

Da regra da multiplicação temos:

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

Exemplo 3: Considerando o Exemplo 1

a. Qual a probabilidade de uma pessoa escolhida ao acaso da população ter olhos azuis dado que possui cabelos loiros?

$$P(A|L) = \frac{P(A \cap L)}{P(L)} = \frac{1768/6800}{2829/6800} = \frac{1768}{2829} = 0,6250$$

Observe que quando condicionamos em L, restringimos o espaço amostral ao conjunto das pessoas loiras. Note que $P(A) = 0,4134 < P(A|L) = 0,6250$ e que os eventos A e L não são independentes pois $P(A|L) \neq P(A)$.

b. Qual a probabilidade de uma pessoa escolhida ao acaso da população não ter cabelos loiros dado que tem olhos castanhos?

$$P(\bar{L}|C) = 1 - P(L|C) = 1 - \frac{115/6800}{857/6800} = 1 - 0,1342 = 0,8658$$

Exemplo 4: Um casal possui 2 filhos sendo que pelo menos um deles é do sexo masculino. Qual é a probabilidade de que ambos sejam do sexo masculino?

Define-se os eventos M ={criança do sexo masculino} e F ={criança do sexo feminino}. Logo, deseja-se obter a probabilidade de que ambos sejam do sexo masculino dado que pelo menos um é do sexo masculino.

$$P(MM|\text{pelo menos um}M) = P(MM)/P(MF \cup FM \cup MM) = (1/4)/(3/4) = 1/3$$

5.2 Avaliação da qualidade de testes diagnósticos

Ao fazer um diagnóstico, um clínico estabelece um conjunto de diagnósticos alternativos com base nos sinais e sintomas do paciente. Progressivamente ele reduz suas alternativas até chegar à uma doença específica.

Alternativamente, ele pode ter fortes evidências de que o paciente tem uma determinada doença e deseja apenas sua confirmação. Para chegar à uma conclusão final o clínico utiliza-se de testes diagnósticos:

- exames de laboratório (ex. dosagem de glicose)
- exame clínico (ex. auscultação do pulmão)
- questionário (ex. CDI (Children's Depression Inventory))

Um teste diagnóstico é um instrumento capaz de diagnosticar a doença com determinada precisão. Para cada teste diagnóstico existe um **valor de referência** que determina a classificação do resultado do teste como **negativo** ou **positivo**.

Um teste diagnóstico é considerado útil quando ele identifica bem a presença da doença. Antes de ser adotado o teste deve ser avaliado para verificar sua capacidade de acerto. Esta avaliação é feita aplicando-se o teste a dois grupos de pessoas: um grupo doente o outro não doente. Nesta fase, o diagnóstico é feito por outro teste chamado padrão ouro.

Os resultados obtidos podem ser organizados de acordo com a tabela abaixo:

Tabela 13: Resultados de um teste para pacientes doentes e não doentes

	Teste		
Doença	+	-	Total
Presente (D)	a	b	a+b
Ausente ($\bar{D}$)	c	d	c+d
Total	a+c	b+d	n

O teste é aplicado a n indivíduos, dos quais sabidamente (a+b) são doentes e (c+d) são não doentes.

Exemplo 5: Em um estudo sobre o teste ergométrico, Wriner et al. (1979) compararam os resultados obtidos entre indivíduos com e sem doença coronariana. O teste foi definido como positivo se foi observado mais de 1mm de depressão ou elevação do segmento ST, por pelo menos 0,08s, em comparação com os resultados obtidos com o paciente em repouso. O diagnóstico definitivo (classificação como doente ou não-doente) foi feito através de angiografia (teste padrão ouro).

Tabela 14: Resultados do teste ergométrico aplicado a 1023 pacientes com doença coronariana e 442 pacientes sem a doença

Doença	Teste Ergométrico					
Coronariana	T_+		T_-		Total	
D_+	815	(a)	208	(b)	1023	(a+b)
D_-	115	(c)	327	(d)	442	(c+d)
Total	930	(a+c)	535	(b+d)	1465	(n)

Sejam os eventos:

- D_+={a pessoa tem doença coronariana}
- D_-={a pessoa não tem doença coronariana}
- T_+={o resultado do teste ergométrico é positivo}
- T_-={o resultado do teste ergométrico é negativo}

Temos interesse em responder duas perguntas:

1. Qual a probabilidade do teste ser positivo dado que o paciente é doente?
2. Qual a probabilidade do teste ser negativo dado que o paciente não é doente?

Em outras palavras, interessa conhecer as probabilidades condicionais:

$$s = P(T_+|D_+) = \frac{P(T_+ \cap D_+)}{P(D_+)} = \frac{a}{a+b}$$

e

$$e = P(T_-|D_-) = \frac{P(T_-|D_-)}{P(D_-)} = \frac{d}{c+d}$$

Estas probabilidades são chamadas **sensibilidade** e **especificidade**. Numa situação ideal a sensibilidade e a especificidade deveriam ser 1.

Exercício: Calcule s e e para o exemplo do teste ergométrico.

Exemplo: Metástase de carcinoma hepático

Lind & Singer (1986) estudaram a qualidade da tomografia computadorizada para o diagnóstico de metástase de carcinoma de fígado, obtendo os resultados sintetizados na Tabela 15. Um total de 150 pacientes foram submetidos a dois exames: a tomografia computadorizada e a laparotomia. Este último é tomado como padrão ouro, isto é, classifica o paciente sem erro.

Metástase de carcinoma hepático	Tomografia computadorizada		Total
	Positivo (T_+)	Negativo (T_-)	
Presente (D_+)	52	15	67
Ausente (D_-)	9	74	83
Total	61	89	150

Tabela 15: Resultados da tomografia computadorizada em 67 pacientes com metástase e 83 sem metástase do carcinoma hepático

A sensibilidade e a especificidade da tomografia são estimados em:

$$s = \frac{52}{67} = 0,776$$

$$e = \frac{74}{83} = 0,892$$

5.3 Valor de predição de um teste

Os índices de sensibilidade e especificidade são ilustrativos e bons sintetizadores das qualidades gerais de um teste mas têm uma limitação séria: **não ajudam a decisão da equipe médica que, recebendo um paciente com resultado positivo do teste, precisa avaliar se o paciente está ou não doente**.

Não se pode depender apenas da sensibilidade e da especificidade, pois estes índices são provenientes de uma situação em que há certeza total sobre o diagnóstico, o que não acontece no consultório médico.

Daí a necessidade destes dois outros índices que refletem melhor a realidade prática.

A ele interessa conhecer o valor de predição positiva (VPP) e o valor de predição negativa (VPN) de um teste:

$VPP = P(D_+|T_+)$ e $VPN = P(D_-|T_-)$

Estes valores são probabilidades condicionais, tal que o evento condicionante é o resultado do teste.

A maneira mais fácil de calcular o VPP e o VPN é através da Tabela 16, sugeriado por Vecchio (1966). Seja $p = P(D_+)$ a prevalência da doença na população de interesse, ou alternativamente a probabilidade de doença pré-teste.

	Proporção com resultado		
População	Positivo	Negativo	Total
Doente	sp	$(1-s)p$	p
Sadia	$(1-e)(1-p)$	$e(1-p)$	$1-p$
Total	$sp+(1-e)(1-p)$	$(1-s)p+e(1-p)$	1

Tabela 16: Probabilidades necessárias para o cálculo dos índices VPP e VPN

O valor de predição positiva é obtido dividindo-se a frequência de pacientes doentes pelo total dos testes positivos.

$$VPP = \frac{sp}{sp+(1-e)(1-p)}$$

De forma análoga, considerando-se os pacientes sadios obtemos o valor de predição negativa

$$VPN = \frac{e(1-p)}{(1-s)p+e(1-p)}$$

- Note que as expressões de VPP e VPN dependem do conhecimento de p.
- Já a sensibilidade e especificidade não dependem do conhecimento de p.

Exemplo: Metástase de carcioma hepático (continuação)

Para uma população cuja prevalência de metástase de carcinoma de fígado é de 2%, o valores de predição da tomografia são:

$$VPP = \frac{0,78 \times 0,02}{0,78 \times 0,02 + (1-0,89)(1-0,02)} = 0,13$$

$$VPN = \frac{0,89 \times (1-0,02)}{0,89(1-0,02) + (1-0,78)0,02} = 0,99$$

Portanto, o VPP é baixo enquanto que o VPN é bastante alto. Se o resultado da tomografia é negativo, a chance de não haver metástase é de 99%. Antes do teste o paciente tinha uma chance de 2% de apresentar a doença, e após o resultado negativo do teste esta chance cai para 1%.

Exercício: Calcule os valores de VPP e VPN para o teste ergométrico.

5.4 Combinação de testes diagnósticos

Muitas vezes, para o diagnóstico de certa doença, dispomos apenas de testes com VPN ou VPP baixos ou se existe um bom teste, este é muito caro ou oferece grande risco e/ou desconforto ao paciente.

Nestas circunstâncias, uma opção é o uso de uma combinação de testes mais simples. A associação de testes eleva a qualidade do diagnóstico, diminuindo o número de resultados incorretos.

Quando dois ou mais testes são usados para se chegar a um diagnóstico é preciso saber como são obtidos os índices de qualidade do testes múltiplos. Restringiremos ao caso de apenas dois testes e as idéias podem ser estendidas para o caso de mais testes.

Formas de combinação de testes

As maneiras mais simples de se formar um teste múltiplo a partir dos resultados de dois testes são os esquemas em paralelo e em série.

No caso do teste em paralelo, se um dos dois testes for positivo o teste conjunto também é positivo.

O teste em série é considerado positivo apenas se os dois testes individuais forem positivos.

Chamando os testes originais de A e B, o teste em paralelo de T_p e o em série de T_s, e usando a linguagem de conjuntos temos que:

$$T_{p+} = A_+ \cup B_+$$

$$T_{s+} = A_+ \cap B_+$$

As sensibilidades e especificidades de T_p e T_s são calculadas com o auxílio das regras de cálculo de probabilidades de eventos vistas anteriormente.

Combinação em paralelo

É de maior utilidade em casos de emergência, quando se necessita de uma abordagem rápida, ou nos casos em que os pacientes são provenientes de lugares distantes.

Teste A	Teste B	Teste em paralelo
-	-	-
-	+	+
+	-	+
+	+	+

Tabela 17: Resultado do teste em paralelo dependendo da classificação dos testes individuais A e B

A sensibilidade do teste em paralelo (s_P) é calculada por:

$$\begin{aligned} s_P &= P(T_{p+}|D_+) \\ &= P(A_+ \cup B_+|D_+) \\ &= P(A_+|D_+) + P(B_+|D_+) - P(A_+ \cap B_+|D_+) \\ &= s_A + s_B - s_A \times s_B \end{aligned}$$

Admitindo-se que o resultado dos dois testes são independentes, especificidade do teste em paralelo (e_p) é:

$$\begin{aligned} e_P &= P(T_{p-}|D_-) \\ &= P(A_- \cap B_-|D_-) \\ &= P(A_-|D_-) \times P(B_-|D_-) \\ &= e_A \times e_B \end{aligned}$$

Os índices VPP e VPN são calculados utilizando-se a sensibilidade e especificidade da combinação em paralelo e a prevalência da doença na população de interesse.

Combinação em série

Nesse caso, os testes são aplicados em consecutivamente, sendo o segundo teste aplicado apenas se o primeiro apresentar resultado positivo. Assim, o custo desse tipo de combinação é menor.

Teste A	Teste B	Teste em série
-	desnecessário	-
+	-	-
+	+	+

Tabela 18: Resultado do teste em série dependendo da classificação dos testes individuais A e B

Se os dois testes A e B são independentes, a sensibilidade (s_s) deste teste é:

$$s_s = P(T_{s+}|D_+) = P(A_+ \cap B_+|D_+) = P(A_+|D_+) \times P(B_+|D_+) = s_A \times s_B$$

A partir de raciocínio análogo, obtemos a expressão para a especificidade (e_s):

$$e_s = e_A + e_B - e_A \times e_B$$

Novamente os valores de VPP e VPN são obtidos usando-se a sensibilidade e especificidade calculadas acima.

Para os cálculos da sensibilidade e especificidade da associação em série e em paralelo, a independência dos dois testes é crucial. Quando os testes não forem independentes, não há uma forma analítica simples para se obter tais índices para um teste composto.

Exemplo: Diagnóstico de câncer pancreático

Imagine um paciente idoso com dores persistentes nas costas e no obdômen e perda de peso. Na ausência de uma explicação para estes sintomas, a possibilidade de câncer do pâncreas é frequentemente levantada. É comum para se verificar esta possibilidade diagnóstica, que ambos os testes de ultra-som e tomografia computdorizada do pâncreas sejam solicitados.

A Tabela 19 apresenta dados hipotéticos sobre os índices s e e dos testes, quando utilizados separadamente e em conjunto (Griner et al., 1981). Note que os esquemas C e D

Teste	Sensibilidade (%)	Especificidade (%)
A: Ultrasom	80	60
B:Tomografia	90	90
C:A ou B positivo	98	54
D:A e B positivo	72	96

Tabela 19: Sensibilidade e especificidade dos testes de ultra-som e tomografia no diagnóstico de câncer de pâncreas

correspodem a testes em paralelo e em série.

Admitindo que o resultado dos dois testes sejam independentes temos as seguintes sensibilidades e especificidades:

$$s_C = 0,8 + 0,9 - 0,8(0,9) = 0,98$$

$$s_D = 0,8(0,9) = 0,72$$

$$e_C = 0,6(0,9) = 0,54$$

$$e_D = 0,6 + 0,9 - 0,6(0,9) = 0,96$$

Quando um ou outro teste é positivo, a sensibilidade combinada é maior que o mais sensível dos testes, mas a especificidade é menor.

Ao contrário, quando o critério para a positividade do teste é que tanto o ultra-som quanto a tomografia seja positivos, a especificidade combinada é maior do que o mais específico dos dois testes separadamente, mas a sensibilidade é menor.

Exemplo: Valores de predição de testes em paralelo e em série

Consideremos dois testes A e B com sensibilidade e especificidade apresentados na Tabela 20 e suponhamos uma prevalência de 1%.

Usando as expressões vistas até agora podemos preencher as outras entradas do corpo da Tabela 20.

Necessidade da combinação de testes

Há pelo menos duas situações em que a necessidade de combinação de testes surge naturalmente: triagem e diagnóstico individual.

Teste	s	e	VPP	VPN
A	0,95	0,900	0,0876	**0,9994**
B	0,80	0,950	0,1391	**0,9979**
Paralelo	**0,99**	0,855	0,0645	**0,9999**
Série	0,76	**0,995**	0,6056	**0,9976**

Tabela 20: Sensibilidade, especificidade e valores de predição de testes individuais A e B e dos testes em série e em paralelo considerando-se uma prevalência de 1%.

Triagem: É um tipo de procedimento que visa classificar pessoas assintomáticas quanto à probabilidade de terem ou não a doença. É aplicado em grande número de pessoas de uma população e não faz por si só o diagnóstico, mas aponta as pessoas que tem maior probabilidade de estarem doentes. Para essas, é indicado um outro teste diagnóstico para comprovar ou não a presença da doença (testes em série).

agnóstico individual: Aparecem dois tipos de combinação: em série e em paralelo. A combinação em paralelo é usada em casos de urgência ou para pacientes residentes em lugares distantes. Uma norma é estabelecida para orientar a decisão do clínico. Por exemplo, se forem realizados 4 testes e 3 ou 4 forem positivos o diagnóstico é de doença.

A combinação em série é usada em consultórios e clínicas hospitalares e em casos de testes caros e arriscados. O paciente realiza primeiramente um teste simples e barato e, se apresentar resultado positivo, é indicado o teste de maior custo ou que oferece riscos ao paciente. Com este procedimento, busca-se reduzir o número de pacientes sadios submetidos a esses testes.

6 Distribuições teóricas de frequências

No início deste material foi feita uma distinção entre variáveis contínuas e discretas.

- Variável discreta: pode assumir qualquer valor dentro de um conjunto enumerável de valores diferentes (ex: número de internamentos diários num certo hospital).
- Variável contínua: pode assumir um valor dentro de um conjunto infinito de valores (ex: peso de um recém-nascido).

Esta distinção é reintroduzida aqui porque a distribuição de dois tipos de variáveis são tratadas de maneiras diferentes em probabilidade. Com variáveis discretas pode-se falar da probabilidade de ocorrência de um valor específico. Com variáveis contínuas, por outro lado, fala-se da probabilidade de se obter um valor dentro de um intervalo específico.

Introduziremos aqui apenas dois modelos probabilísticos usados para variáveis discretas e contínuas.

6.1 A distribuição Binomial

A distribuição binomial é um exemplo de uma distribuição discreta. Ela lida com situações em que cada resultado de uma série de ensaios independentes resulta num dentre dois resultados possíveis.

Mais formalmente, considere um ensaio realizado n vezes, sob as mesmas condições, com as seguintes características:

1. cada repetição do ensaio produz um dentre dois resultados possíveis, denominados tecnicamente por sucesso (S) ou fracasso (F), ou seja, a resposta é dicotômica.
2. a probabilidade de sucesso, $P(S) = p$, é a mesma em cada repetição do experimento. (Note que $P(F) = 1 - p$).
3. os ensaios são independentes, ou seja, o resultado de um ensaio não interfere no resultado do outro.

O número total de sucessos X é uma variável aleatória com distribuição binomial com parâmetros n e p e é por denotada $X \sim B(n, p)$.

De maneira genérica, a probabilidade de $X = x$, pode ser encontrada como:

$$P(X = x) \;=\; \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x = \{0, 1, 2, \cdots\}. \tag{5}$$

A **média** de um variável aleatória binomial é np e a **variância** é $np(1-p)$.

6.1.1 Exemplo: A arte da apreciação de vinho.

Suponha que estejamos interessados em estudar a arte da apreciação de vinho. Como parte de nossos estudos, pedimos para uma pessoa (Pierre), que se diz um connoisseur de vinhos, provar duas taças de vinho e escolher o mais caro.

Esta tarefa é repetida 10 vezes, cada vez com um par diferente de vinhos.

Assuma a princípio que o apreciador não sabe nada sobre vinhos e que ele na verdade escolhe os vinhos ao acaso e que está simplesmente tentando esconder este fato.

Assumindo que não existem fatores externos que interferem na decisão do apreciador (tal como uma tendência a escolher o vinho mais escuro), então em cada ensaio a probabilidade dele acertar na escolha é de 0,5.

Agora suponha que o apreciador tenha acertado 9 dos 10 ensaios.

Pergunta: Qual é a probabilidade de 9 acertos em 10? Qual a sua opinião a respeito da habilidade daquela pessoa em reconhecer um bom vinho?

Seja X o número de acertos em 10 ensaios, então desejamos saber $P(X = 9)$.

Um possível resultado que satisfaz o critério $X = 9$ seria, nove escolhas corretas seguida de uma incorreta: $C, C, C, C, C, C, C, C, I$. A probabilidade deste resultado (pela regra da multiplicação) é a probabilidade conjunta:

$$P(C, C, C, C, C, C, C, C, I) = (0, 5^9)(0, 5^1) = p^x(1 - p)^{(n-x)}$$

(Lembre-se que estamos assumindo independência entre os ensaios).

Outras combinações levam ao mesmo resultado de respostas 9 corretas em 10 ensaios (por ex,$C, C, C, C, C, C, C, I, C$), cada uma com a mesma probabilidade $(0, 5^9)(0, 5^1)$. Na verdade, um total de 10 combinações diferentes levarão a 9 escolhas corretas em 10 ensaios, ou seja, $C_x^n = C_9^{10} = 10$.

Assim, pela regra da adição, temos que a probabilidade de 9 acertos em 10 ensaios é de:

$$P(X = 9) = C_9^{10}(0, 5^9)(0, 5^1) = 10(0, 001953)(0, 5) = 0, 0098.$$

Então, a probabilidade de fazer 9 escolhas corretas em 10 ensaios com $p = 0, 5$ é remota, ocorrendo aproximadamente 1 vez a cada 100 sessões.

Este resultado nos leva ao seguinte questionamento: Será que temos evidência suficiente para rejeitar nossa conjectura inicial de que Pierre não sabe nada sobre vinhos?

6.1.2 Exercício: Albinismo em humanos.

Suponha que num pedigree humano envolvendo albinismo (o qual é recessivo), nós encontremos um casamento no qual sabe-se que ambos os parceiros são heterozigotos para o gene albino. De acordo com a teoria Mendeliana, a probabilidade de que um filho desse casal seja albino é um quarto. (Então a probabilidade de não ser albino é $\frac{3}{4}$.)

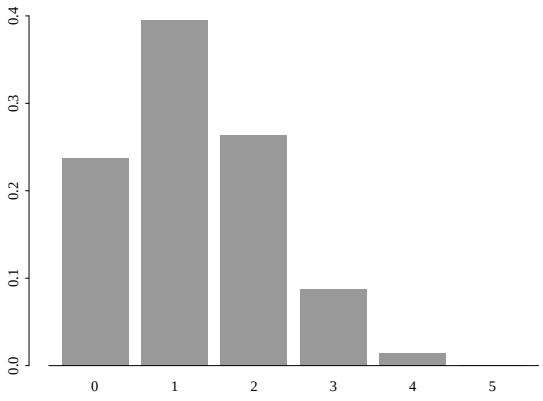
Agora considere o mesmo casal com 2 crianças. A chance de que ambas sejam albinas é $(\frac{1}{4})^2 = \frac{1}{16} = 0.0625$. Da mesma forma, a chance de ambas serem normais é $(\frac{3}{4})^2 = \frac{9}{16} = 0.5625$. Portanto, a probabilidade de que somente uma seja um albina deve ser $1 - \frac{1}{16} - \frac{9}{16} = \frac{6}{16} = \frac{3}{8} = 0.375$.

Alternativamente, poderiamos ter usado a formula acima definindo como variável aleatória X o número de crianças albinas, com $n = 2$, $p = \frac{1}{4}$, e estariamos interessados em $P(X = 1)$.

Se agora considerarmos a família com $n = 5$ crianças, as probabilidades de existam $x = 0, 1, 2, \ldots, 5$ crianças albinas, em que a probabilidade de albinismo é $p = \frac{1}{4}$, são dadas por

$$P(X = x) = \frac{5!}{x!(5-x)!} \left(\frac{1}{4}\right)^x \left(\frac{3}{4}\right)^{5-x} \tag{6}$$

as quais ficam como segue.



O número esperado (ou média) de crianças albinas em famílias com 5 crianças para casais heterozigotos para o gene albino é $np = 5 \times \frac{1}{4} = 1,25$.

6.2 A distribuição Normal

A **distribuição Normal** é a mais familiar das distribuições de probabilidade e também uma das mais importantes em estatística.

Exemplo: O peso de recém-nascidos é uma variável aleatória contínua. A Figura 32 e Figura 33 abaixo mostram a distribuição de frequências relativas de 100 e 5000 pesos de recém-nascidos com intervalos de classe de 500g e 125g, respectivamente.

O segundo histograma é um refinamento do primeiro, obtido aumentando-se o tamanho da amostra e reduzindo-se a amplitude dos intervalos de classe. Ele sugere a curva na Figura 34, que é conhecida como curva normal ou Gaussiana.

A variável aleatória considerada neste exemplo e muitas outras variáveis da área biológica podem ser descritas pelo modelo normal ou Gaussiano.

A equação da curva Normal é especificada usando 2 parâmetros: a **média** μ, e o **desvio padrão** σ.

Denotamos $\mathrm{N}(\mu, \sigma)$ à curva Normal com média μ e desvio padrão σ.

A média refere-se ao centro da distribuição e o desvio padrão ao espalhamento (ou achatamento) da curva.

A distribuição normal é simétrica em torno da média o que implica que a média, a mediana e a moda são todas coincidentes.

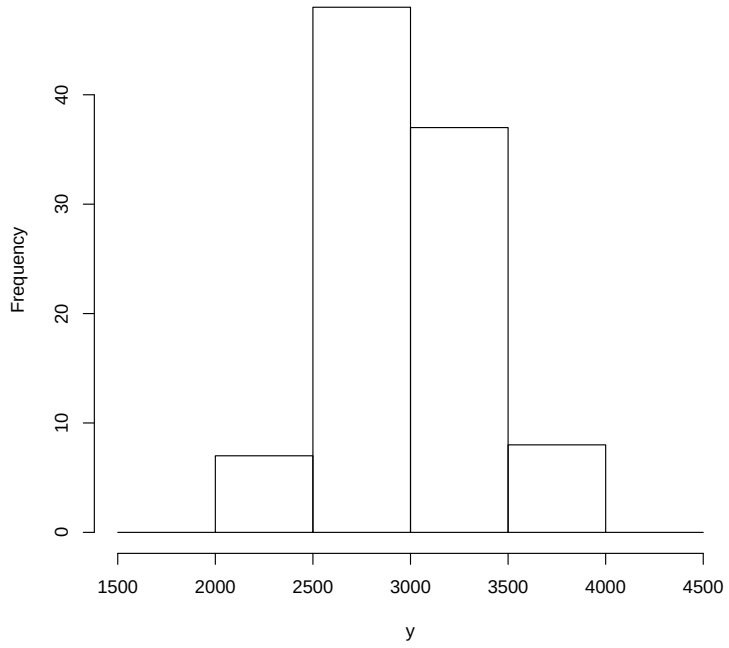


Figura 32: Histograma de frequências relativas a 100 pesos de reçém-nascidos com intervalo de classe de 500g

Para referência, a equação da curva é

$$f(x) = \frac{1}{\sqrt{(2\pi\sigma^2)}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \tag{7}$$

Felizmente, você não tem que memorizar esta equação. O importante é que você entenda como a curva é afetada pelos valores numéricos de μ e σ. Isto é mostrado no diagrama da Figura 35.

A área sob a curva normal (na verdade abaixo de qualquer função de densidade de probabilidade) é 1. Então, para quaisquer dois valores específicos podemos determinar a proporção de área sob a curva entre esses dois valores.

Para a distribuição Normal, a proporção de valores caindo dentro de um, dois, ou três desvios padrão da média são:

Range	Proportion
$\mu \pm 1\sigma$	68.3%
$\mu \pm 2\sigma$	95.5%
$\mu \pm 3\sigma$	99.7%

Exemplo: Suponhamos que no exemplo do peso do recém-nascidos $\mu = 2800g$ e $\sigma = 500g$. Então:

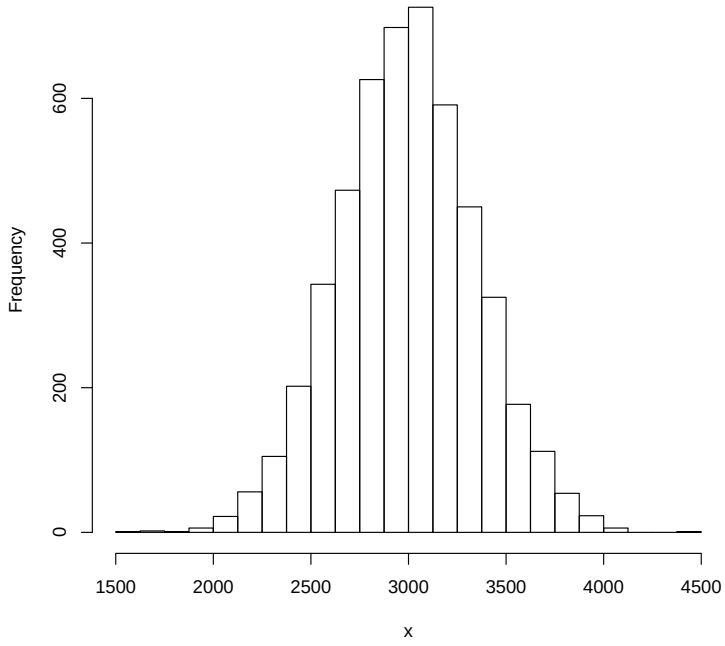


Figura 33: Histograma de frequências relativas a 5000 pesos de reçém-nascidos com intervalo de classe de 125g

$P(2300 \leq X \leq 3300) = 0,683$
$P(1800 \leq X \leq 3800) = 0,955$
$P(1300 \leq X \leq 4300) = 0,997$

Usando este modelo podemos dizer que cerca de 68% dos recém-nascidos pesam entre 2300g e 3300g. O peso de aproximadamente 95% dos recém-nascidos está entre 1800g e 3800g. Praticamente todos os bebês desta população nascem com peso no intervalo (1300,4300).

Na prática desejamos calcular probabilidades para diferentes valores de μ e σ.

Para isso, a variável X cuja distribuição é $N(\mu, \sigma)$ é transformada numa forma padronizada Z com distribuição $N(0, 1)$ (**distribuição normal padrão**) pois tal distribuição é tabelada.

A quantidade Z é dada por

$$Z = \frac{X - \mu}{\sigma} \tag{8}$$

Exemplo: Suponha que a pressão arterial sistólica em pessoas jovens saudáveis tenha distribuição $N(120, 10)$.

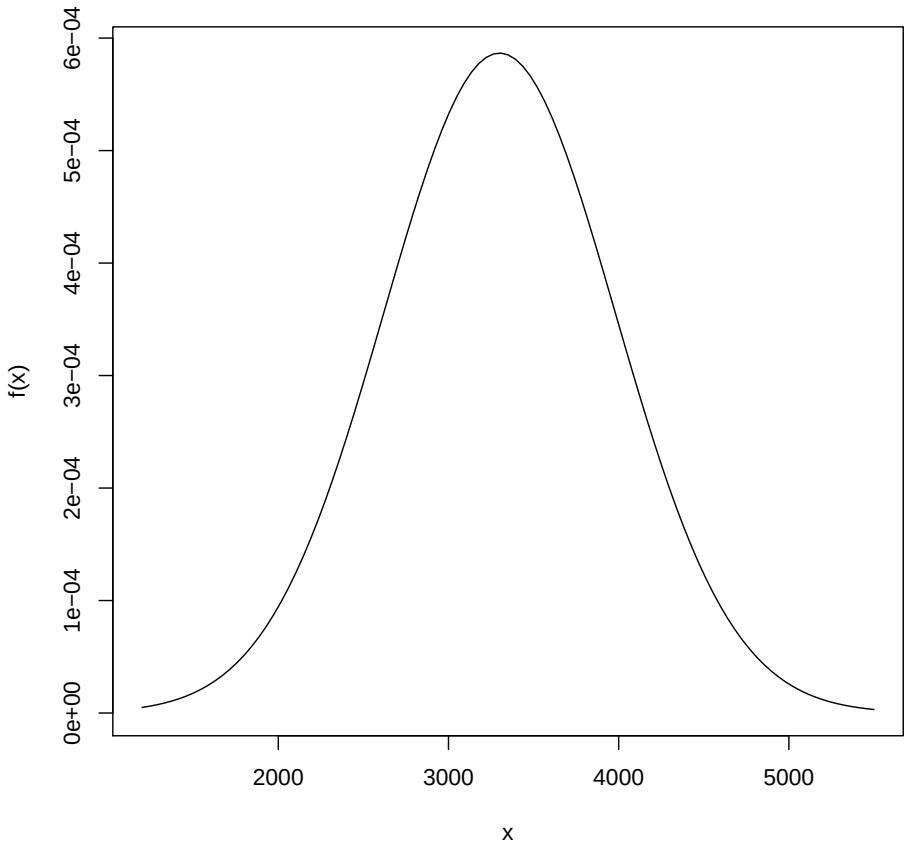


Figura 34: Função de densidade de probabilidade para a variável aleatória contínua X=peso do recém-nascido (g)

1. Qual é a probabilidade de se encontrar uma pessoa com pressão sistólica acima de $140 mmHg$?

$$\begin{aligned} P(X \geq 140) &= P(\tfrac{X-120}{10} \geq \tfrac{140-120}{10}) = \\ &= P(Z \geq 2) = \\ &= 1 - P(Z < 2) = \\ &= 1 - 0,9772 = \\ &= 0,0228 \end{aligned}$$

   Ou seja, 2,28% das pessoas jovens e sadias têm pressão sistólica acima de 140 $mmHg$.

2. Quais são os limites de um intervalo simétrico em relação à média que engloba 95% dos valores das pressões sistólicas de pessoas jovens e sadias?

   Neste caso queremos encontrar a e b tais que: $P(a \leq X \leq b) = 0,95$.

   Primeiro padronizamos essa probabilidade, isto é,

$$P(a \leq X \leq b) = P\left(\frac{a-120}{10} \leq \frac{X-120}{10} \leq \frac{b-120}{10}\right) = 0,95$$

   Ou seja, $P(a' \leq Z \leq b') = 0,95$. Escolhendo uma solução simétrica temos $-a' = b'$.

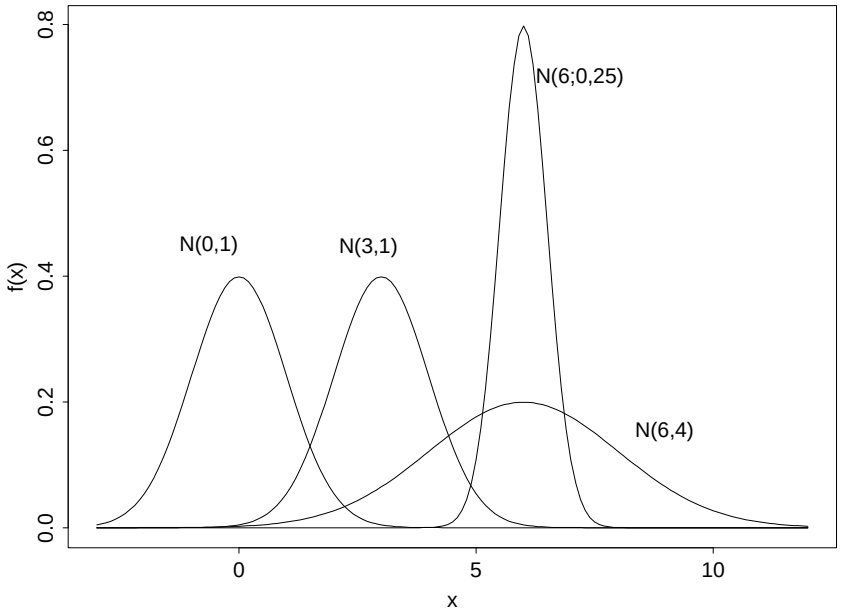


Figura 35: distribuições normais com mesma média μ e vários valores de σ

Como $P(Z \leq b') = 0.975$, da tabela da gaussiana padrão obtemos $a' = -1,96$ e $b' = 1,96$. Consequentemente, $\frac{a-120}{10} = -1,96$ e $\frac{b-120}{10} = 1,96$, ou seja, $a = 100,4$ e $b = 139,6$.
Para essa população de jovens saudáveis, o intervalo $[100,4; 139,6]$ engloba 95% dos valores pressóricos, isto é, aproximadamente entre 100 $mmHg$ e 140 $mmHg$.

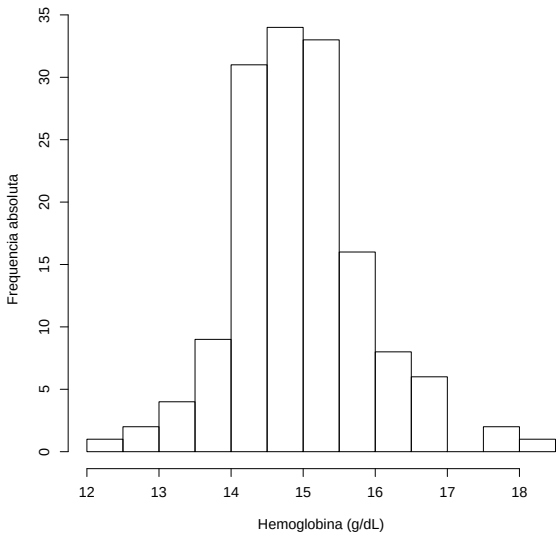


Figura 36: Histograma da hemoglobina de mulheres sadias

7 Construção de faixas de referência

Trataremos aqui da construção de faixas de referência ou simplesmente de um valor de referência.

Tal procedimento permite a caracterização do que é típico em uma determinada população. É empregado largamente em Ciências da Saúde, por exemplo, nos resultados de exames de laboratório. Entretanto, tal metodologia tem muitas outras aplicações, tais como a determinação de níveis toleráveis de barulho ou a caracterização dos níveis de poluição em uma região.

7.1 Conceito de faixa de referência

O histograma mostrado na Figura 36 trata-se de taxas de hemoglobina (g/dl) de 147 mulheres clinicamente sadias.

Tomando o histograma desta figura como aproximação para a sua distribuição entre, vemos que valores menores que 12 g/dL são pouco comuns. É, pois, razoável considerar um diagnóstico de anemia ao se observar uma paciente com valor de hemoglobina de 11 g/dL.

Por causa da variabilidade entre indivíduos, a análise de um valor específico de hemoglobina em mulheres deve ser feita confrontando-o com a faixa de 12-16 g/dL. É a chamada *faixa de referência*.

Resumindo, construiu-se uma faixa de referência para a taxa de hemoglobina em mulheres, tendo por base um grande número de mulheres sadias e estudando a forma da distribuição dos dados.

O procedimento sugerido acima é geral. Todo exame de laboratório que produz uma medida como resultado é analisado confrontando-se seu valor com uma faixa de referência.

Hipótese de construção:
A construção de faixas de referência baseia-se na hipótese de que a população de sadios e a de

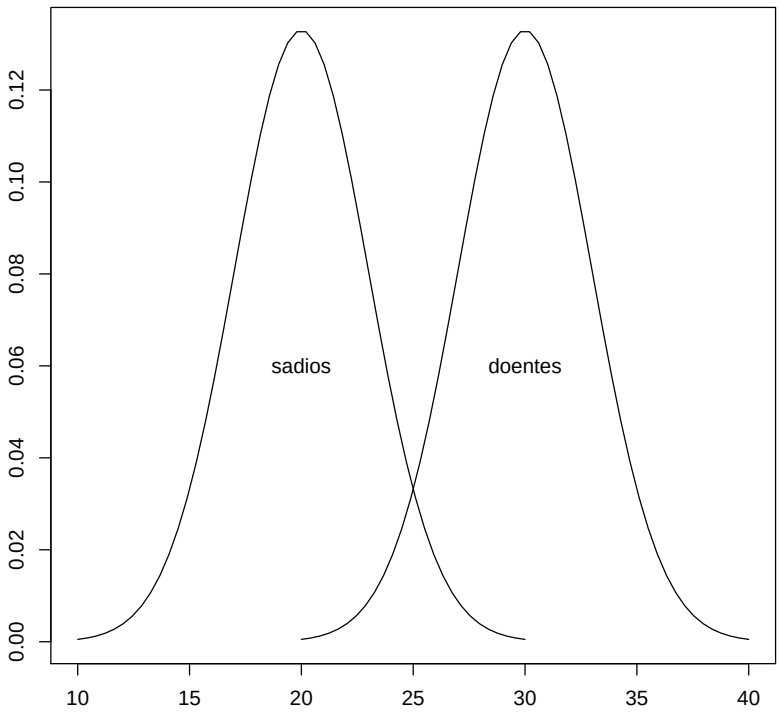


Figura 37: Modelo para construção de faixas de referência

doentes produzem para a medida de interesse valores que flutuam em torno de médias diferentes, gerando curvas com alguma interseção. A Figura 37 ilustra a situação quando as distribuições de sadios e doentes são Gaussianas.

Veremos dois métodos para construção de faixas de referência: o **método da curva de Gauss** e o **método dos percentis**.

7.2 Método da curva de Gauss

Este método pressupõe que a variável de interesse tem distribuição Gaussiana (normal). Portanto, antes de utilizá-lo, é necessário verificar se as observações dos indivíduos sadios provém de uma distribuição normal ou aproximadamente normal.

Uma faixa de referência usual considera aproximadamente 95% dos indivíduos sadios, cujos limites, conforme vimos são:

$$\mu \pm 2\sigma$$

De um modo geral, μ e σ são desconhecidos, mas para a construção dessas faixas devemos nos basear num grande número de indivíduos sadios. Assim, é de se esperar que tanto $\bar{x}$ quanto

s estejam próximos de μ e σ. Consequentemente, as faixas de normalidade são construídas a partir das informações amostrais, ie:

$$\bar{x} \pm 2s$$

De maneira análoga, podem ser obtidas outras faixas de referência compreendendo outras porcentagens de indivíduos sadios, tais como, 90%, 98%, etc.

Exemplo: Sabendo-se que a taxa de hemoglobina (g%) em mulheres sadias tem distribuição $N(12, 2)$, construa as faixas de referência que englobem:

1. 95% das taxas de hemoglobina
2. 90% das taxas de hemoglobina

Exemplo: Embora gastroenterologistas infantis reconhecessem a utilidade diagnóstica do teor de gordura fecal, até 1984 não existia um padrão de referência desta medida para crianças brasileiras. Para preencher esta lacuna, o prof Francisco Penna, titular de pediatria da UFMG, examinou 43 crianças sadias que produziram os valores da Tabela 21, expressos em g/24 hs.

3,7	1,6	2,5	3,0	3,9	1,9	3,8	1,5	1,1
1,8	1,4	2,7	2,1	3,3	3,2	2,3	2,3	2,4
0,8	3,1	1,8	1,0	2,0	2,0	2,9	3,2	1,9
1,6	2,9	2,0	1,0	2,7	3,0	1,3	1,5	4,6
2,4	2,1	1,3	2,7	2,1	2,8	1,9		

Tabela 21: Teor de gordura fecal (g/24hs) de 43 crianças sadias

A média e o desvio-padrão dos dados são: $\bar{x} = 2,303; s = 0,872$. Utilizando o método da curva de Gauss obtemos a seguinte faixa de referência com aproximadamente 95% de cobertura: $(2,303 - 2 \times 0,872; 2,303 + 2 \times 0,872) = (0,559; 4,047)$. Como neste caso só interessa o limite superior, o valor de referência que é superado por apenas 5% dos teores de gordura fecal é calculado por: $2,303 + 1,64 \times 0,872 = 3,733$.

7.3 Método dos percentis

Um método alternativo para obter valores de referência é o método dos percentis. Este não exige qualquer suposição sobre a forma da distribuição. Este método pode ser utilizado para a situação em que os dados estão ou não agrupados.

A idéia é determinar uma faixa que concentre um determinado percentual da população. Por exemplo, se fixarmos um percentual de 95%, a construção da faixa de referência consistiria em determinar o percentil de ordem 2,5 e 97,5.

Quando os dados estão agrupados, pode-se aplicar o método dos percentis utilizando-se a ogiva.

Exemplo: Teor de gordura fecal em crianças sadias (continuação)

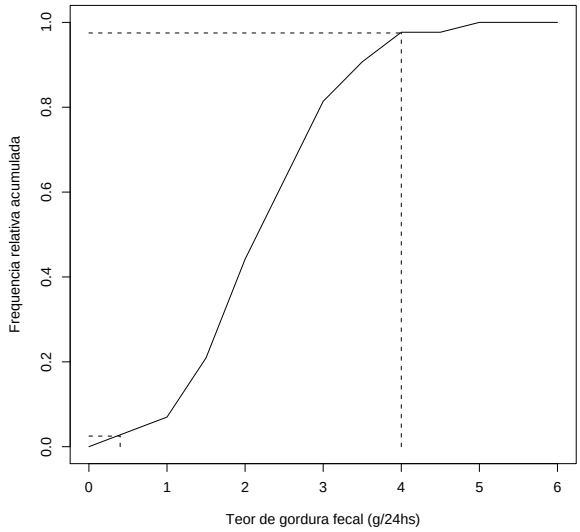


Figura 38: Ogiva do teor de gordura fecal de criaças sadias.

Pode-se obter a faixa de referência de 95% projetando-se os percentis de ordem 0,025 e 0,975, que são 0,4 g/24hs e 4 g/24hs (Figura 38). Portanto, espera-se que o teor de gordura fecal de 95% das crianças sadias da população estudada esteja entre esses limítrofes. Em termos do limite superior de 95%, teriamos um valor de referência de 3,8 g/24hs.

Para dados não agrupados, o método consiste em ordenar os valores da amostra de 1 a n, e diante de cada um colocar o valor de $(i - 0,5)/n$, em que i é o número de ordem da observação. Este valor é uma boa aproximação para a ordem do percentil, correspondente ao dado de ordem i.

7.4 Exemplo: Rastreamento de doadores de sangue

A dosagem sérica da enzima alanina transaminase (ALT), é usada desde 1955 como indicadora de danos hepatocelulares, encontrando-se pronunciadamente aumentada nos casos de hepatites virais, agudas e crônicas, sendo mais específica para detectar doença hepática assintomática em pacientes não alcoólicos.

Nos bancos de sangue, os doadores são rotineiramente avaliados para a elevação dessa enzima antes da transfusão.

O exemplo a seguir trata da transmissão da hepatite virótica na transfusão de sangue (Colton, 1974). Na época foi implantado um programa de rastreamento de sangue contaminado com vírus da hepatite em doadores de um banco de sangue, baseado na medida de ALT.

Resultados empíricos de várias pesquisas indicam que o logaritmo da determinação do ALT tem aproximadamente distribuição gaussiana nas populações sem e com danos hepatocelulares, respectivamente denominadas, *sadia* e *doente*.

Os parâmetros destas populações na escala logaritmica em UI/L estão mostrados no quadro

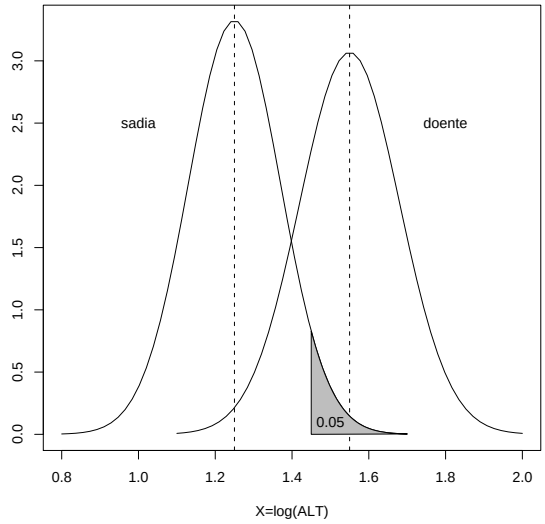


Figura 39: Distribuição da população sadia e doente.

abaixo:

População	Média	Desvio-padrão
sadia	1,25	0,12
doente	1,55	0,13

Para definir o rastreamento proposto, foi necessário estbeecer um ponto de corte de $X = log(ALT)$.

Assim, o sangue de um doador em potencial com x acima deste ponto de corte seria rejeitado e abaixo seria aceito.

Foi estabelecido que este ponto de corte seria tal que o procedimento de rastreamento aceitasse 95% dos doadores provenientes da população sadia.

1. Como determinamos este ponto de corte (x)?
2. Qual é o percentual de amostras aceitas indevidamente?
3. É possível calcular a especificidade deste teste? Se sim, qual é a sua especificidade?
4. É possível calcular a sensibilidade deste teste? Se sim, qual é a sensibilidade do teste?
5. Suponha que a prevalência de danos hepatocelulares neste população seja de 12%. Quais os valores preditivos deste teste?

   Para responder esta pergunta podemos usar as fórmulas ou alternativamente podemos simular uma população de determinado tamanho e construir uma tabela 2 x 2 a aprtir dos percentuais fixados. Por exemplo, numa população de 1000 indivíduos, seriam 120 doentes (ou seja, 12% de 1000) e 880 sadios.

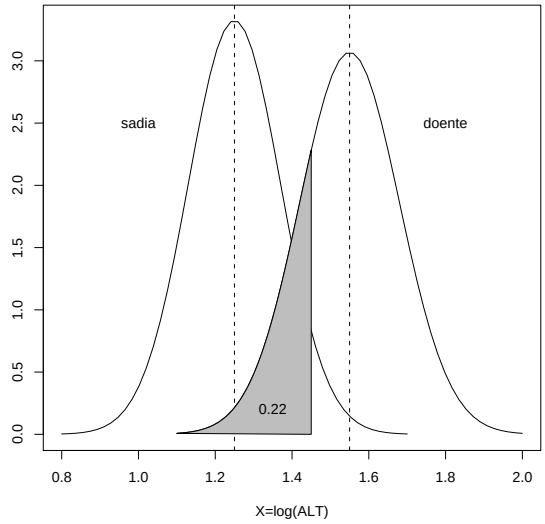


Figura 40: Distribuição da população sadia e doente.

Considerando o ponto de corte de 1,45 e portanto os percentuais de amostras aceitas e rejeitadas indevidamente seriam de 22,06% e 5%, espera-se que 26 (22,06% de 120) doadores doentes sejam aceitos e 44 (5% de 880) doadores saudáveis sejam rejeitados.

Doador proveniente de população	Amostra		
	Aceita	Rejeitada	Total
Sadia	836	44	880
Doente	26	94	120
Total	862	138	1000

Tabela 22: Distribuição de amostras de sangue aceitas e rejeitadas numa população de 1000 indivíduos

A efetividade do rastreamento pode ser avaliada através dos valores preditivos.

$$VPP = 94/138 = 0,68(68\%)$$

$$VPN = 836/862 = 0,97(97\%)$$

Em resumo, usando o ponto de corte de 1,45 tem-se um alto percentual de amostras de doadores sadios sendo rejeitadas (32%).

6. De que modo seria possível diminuir o percentual de amostras rejeitadas indevidamente?

   Se, em uma tentativa de minimizar a rejeição de sangue em doadores saudáveis, passarmos a aceitar 97,5% dos doadores originados da população sadia, o ponto de corte seria 1,49.

   Nesse caso, apenas 2,5% de sangue seriam rejeitados indevidamente, mas o percentual de doentes com sangue aceito passaria para 32,3%.

7.5 Considerações finais

1. Um indivíduo com resultado fora da faixa de referência deve ser considerado um paciente que necessita de mais investigações.

2. O tamanho da amostra é crucial na obtenção de faixas de referência representativas. Tal representatividade depende da escolha adequada do método de contrução e da variabilidade dos dados.

3. Podem existir indivíduos sadios fora da faixa de referência e indivíduos doentes cuja medida pertence à faixa de referência.

8 Medida do efeito de uma intervenção ou exposição

Nos estudos do tipo caso-controle, é de interesse verificar se o grupo de pacientes com a patologia de interesse foi mais (ou menos) exposto ao fator de risco em análise.

Podemos verificar a *existência* de efeito da exposição através de testes de hipóteses, assunto que será visto na próxima seção. No entanto, com os testes de hipóteses, damos uma resposta parcial a essa questão.

Vamos nesta seção, responder esta pergunta **medindo ou quantificando** o efeito da exposição.

Antes porém, alguns resultados gerais serão introduzidos.

8.1 Conceitos fundamentais

Para apresentar os conceitos básicos necessários para compreensão deste material, vamos utilizar um exemplo em que existe apenas uma variável resposta de interesse, descrita pela distribuição gaussiana.

Exemplo: Níveis plasmáticos de vitamina A

Moura (1990) avaliou os níveis plasmáticos de vitamina A num grupo de 41 crianças diabéticas com idade de até 12 anos.

Um de seus interesses era conhecer o nível sanguíneo de vitamina A neste grupo, composto de pacientes típicos dos atendidos pelo setor de Endocrinologia Pediátrica da Faculdade de Medicina da UFMG.

8.1.1 Parâmetro de interesse

Em termos estatísticos o objetivo deste estudo é conhecer o nível médio, μ, de vitamina A no sangue de crianças diabéticas. Neste caso, μ é o *parâmetro* de interesse.

Em geral, para a análise estatística dos dados de um problema clínico é preciso:

1. identificar os parâmetros de interesse, isto é, aquelas características numéricas cujo conhecimento viabiliza a solução da questão colocada.
2. tomar uma decisão baseada no valor do parâmetro de interesse que, entretanto **não é conhecido** na prática.

Então, utilizando a teoria da estimação aprendemos como, a partir de dados amostrais, podemos obter valores que se aproximam do verdadeiro valor do parâmetro.

8.1.2 Estimadores

É razoável admitir que o nível sanguíneo de vitamina A tem distribuição gaussiana.

Com esta especificação, o problema se reduz a encontrar estimativas para os dois parâmetros da distribuição (μ, σ).

A teoria estatística tem várias soluções para este problema. Uma possível solução é dada pelo *método da máxima verossimilhança*.

Intuitivamente, este método baseia-se no seguinte princípio:

"entre todos os valores possíveis para os parâmetros (μ, σ), escolhemos aqueles que são **mais prováveis** de terem produzido o conjunto de dados observado."

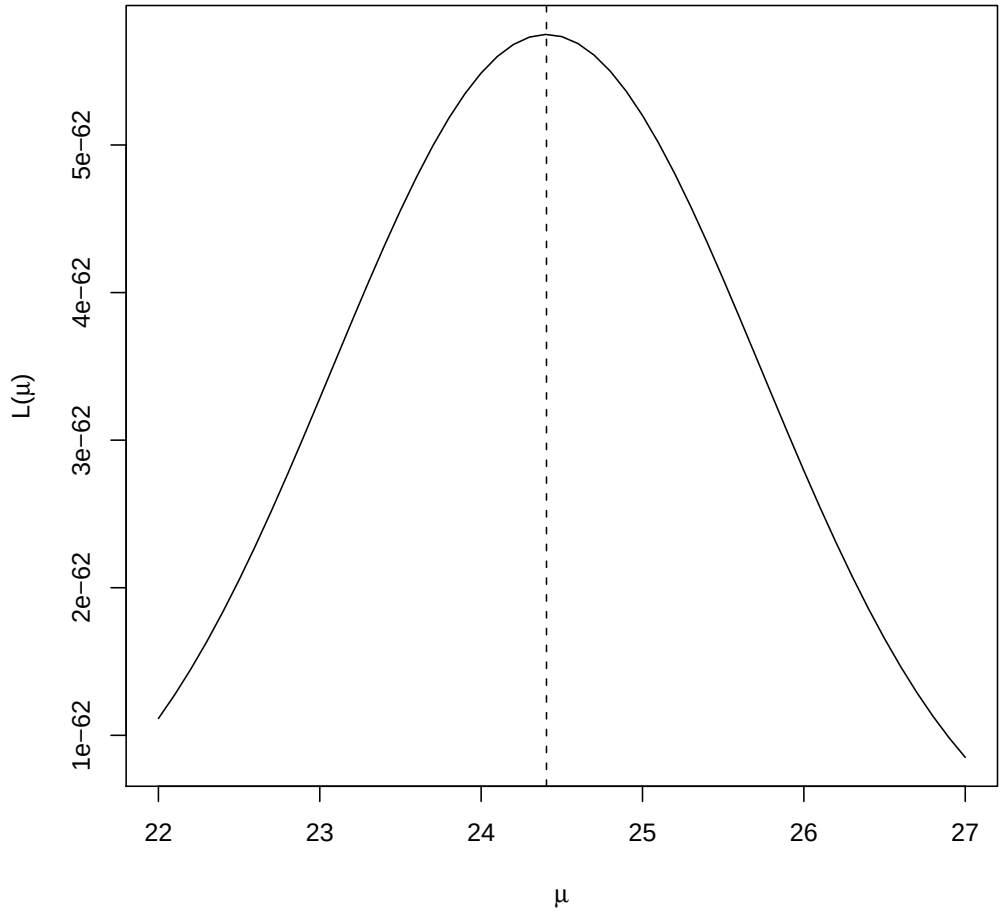


Figura 41: Função de verossimilhança para μ assumindo σ conhecido.

Assumindo que os níveis de vitamina A são representados por $X_1, X_2, \cdots, X_n$ $(n = 41)$, e representando por $\hat{\mu}$ e $\hat{\sigma}^2$ os estimadores fornecidos pelo método da máxima verossimilhança, temos que

$$\hat{\mu} = \bar{X}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

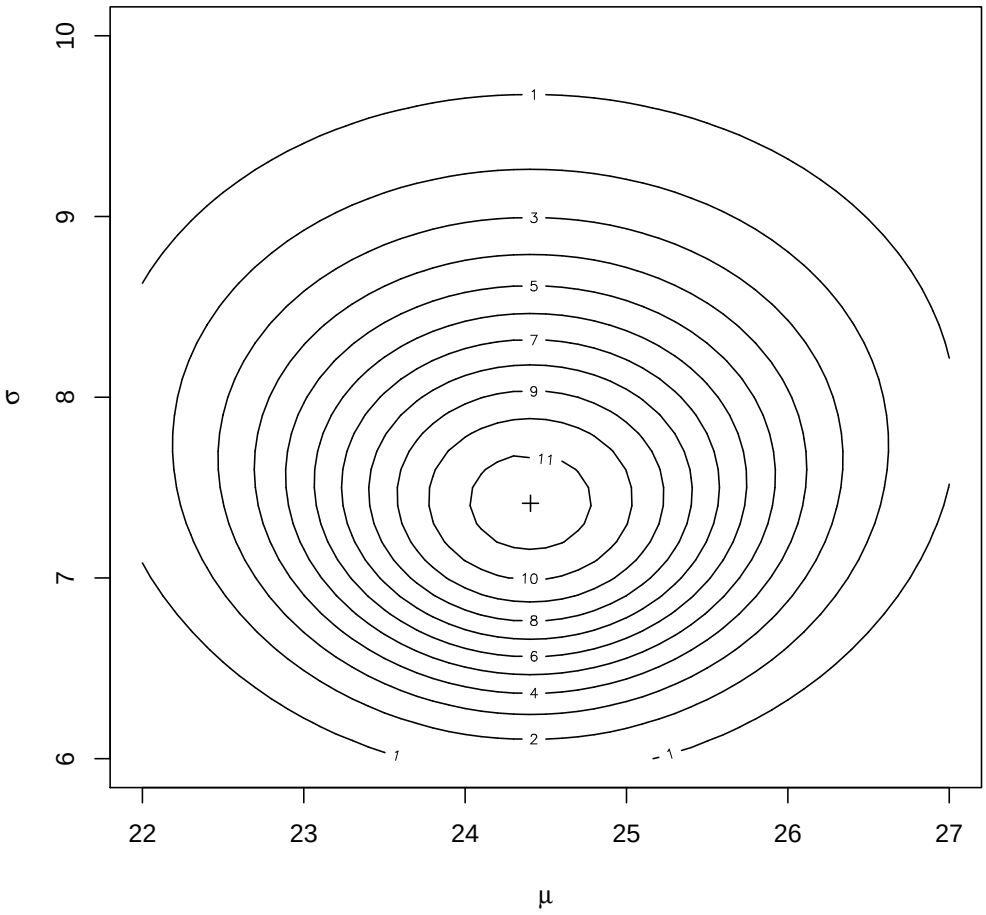


Figura 42: Contornos de verossimilhança para μ e σ.

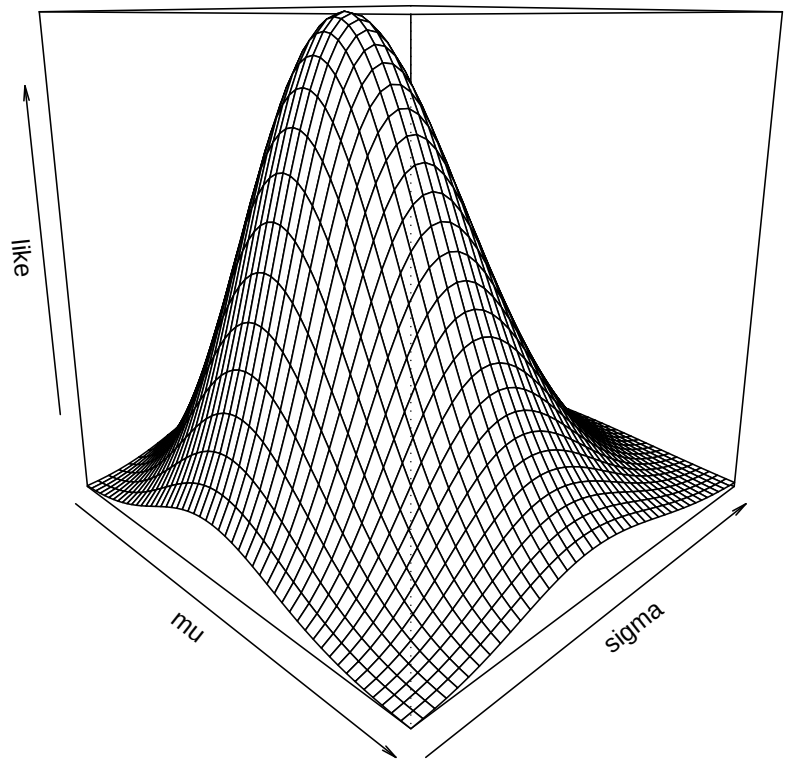


Figura 43: Superfície de verossimilhança para μ e σ.

Recomendamos, entretanto, para σ^2 o seguinte estimador:

$$\hat{S}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1}$$

Existem motivos teóricos e práticos para esta indicação que fogem ao nível teórico pretendido neste material.

Estes dois estimadores são chamados de **pontuais**, pois fornecem apenas um único valor de estimativa do parâmetro.

A questão da sua variabilidade é tratada pela construção de um **intervalo de confiança**.

8.1.3 Intervalo de confiança

Diferentes amostras podem ser retiradas de uma mesma população, e amostras diferentes podem resultar em estimativas diferentes. Isto é, um estimador é uma variável aleatória, podendo assumir valores diferentes para cada amostra.

Então, ao invés de estimar o parâmetro de interesse por um único valor, é muito mais informativo estimá-lo por um intervalo de valores que considere a variação presente na amostra e que contenha o seu verdadeiro valor com determinada confiança.

Este intervalo é chamado de **intervalo de confiança**.

Para construir um intervalo de confiança precisamos conhecer a distribuição de probabilidade do estimador. Lembre que um estimador é uma variável aleatória e que uma variável aleatória é completamente caracterizada por sua distribuição de probabilidade.

Por exemplo, o intervalo de confiança para μ, é obtido usando o seguinte resultado da Teoria Estatística:

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

Então, mesmo não se conhecendo μ, podemos calcular probabilidades envolvendo a estatística T.

Isso nos permite construir o *intervalo de confiança* para μ da seguinte forma:

Através da tabela da distribuição t_{n-1} obtemos um valor t^* tal que

$$P(-t^* \leq t_{n-1} \leq t^*) = 1 - \alpha$$

ou seja,

$$P(-t^* \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t^*) = 1 - \alpha$$

que pode ser reescrita como

$$P(\bar{X} - t^*\frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + t^*\frac{S}{\sqrt{n}}) = 1 - \alpha$$

Assim, um intervalo de confiança para μ com nível de confiança $100(1-\alpha)\%$ é dado por

$$\left[\bar{X} - t^* \frac{S}{\sqrt{n}}; \bar{X} + t^* \frac{S}{\sqrt{n}}\right]$$

Portanto, o intervalo de confiança para μ é centrado na estimativa do efeito, e varia de uma quantidade t^* desvios-padrão para baixo até o mesmo número de desvios-padrão para cima.

Exemplo: Níveis plasmáticos de vitamina A (continuação)

Para os dados deste estudo, $\bar{x} = 25,5mcg/dL$ e $s = 8,5mcg/dL$.

Para um nível de confiança de 95%, o percentil da distribuição com 40 graus de liberdade é $t^*_{0,975} = 2,021$.

Portanto, o intervalo de confiança para a média do nível plasmático de vitamina A é $[22,8; 28,2]$.

Com 95% de confiança podemos afirmar que o nível médio plasmático de vitamina A em crianças diabéticas com idade de até 12 anos varia entre $22,8mcg/dL$ e $28,2mcg/dL$.

8.1.4 Distribuição da média amostral

Nesta seção serão apresentados resultados da Teoria Estatística, usados na construção de intervalos de confiança.

São três os resultados a serem apresentados:

1. a fórmula do desvio-padrão da média,
2. a distribuição da média amostral e
3. sua distribuição padronizada pelo desvio-padrão.

Consideremos que a variável aleatória nível de hemoglobina em mulheres jovens e sadias tenha distribuição gaussiana com média 12 e desvio-padrão 1.

Vamos tirar sucessivas amostras de tamanho 6 desta população e para cada amostra, calcular sua média.

Para 50 repetições, obtivemos os resultados da Tabela 23.

Resultados:

1. O desvio-padrão da média é dado pelo desvio-padrão da população dividido por $\sqrt{n}$, ou seja, $\sigma_{\bar{x}} = \sigma/\sqrt{n}$;

   Neste exemplo, deveríamos obter $\frac{1}{\sqrt{6}} = 0,41$. A diferença entre o valor teórico e o valor observado 0,359 advém do fato de estarmos usando uma amostra das médias.

Amostra	x_1	x_2	x_3	x_4	x_5	x_6	Média
1	11,7	11,3	11,3	11,8	11,7	12,7	11,8
2	12,9	12,5	9,9	11,7	12,9	12,3	12,0
3	12,3	13,5	12,8	12,6	10,9	12,8	12,5
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
50	11,2	11,4	12,1	13,4	11,9	11,5	11,9
						Média	12,072
						desvio-padrão	0,359

Tabela 23: Amostras de tamanho 6 e suas médias

2. O Teorema Central do Limite, diz que a média tem aproximadamente distribuição gaussiana centrada na média da população e desvio-padrão dado por $\sigma_{\bar{x}} = \sigma/\sqrt{n}$, isto é,

$$\bar{X} \sim N(\mu, \sigma/\sqrt{n})$$

3. O terceiro resultado é devido a W. Gosset que descobriu, que a razão

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tem distribuição que não depende de μ e σ^2, mas apenas do tamanho da amostra (a distribuição t de Student).

8.2 Medida do efeito: resposta contínua

Veremos aqui como o efeito pode ser estimado a partir dos dados e como deve ser tratada a questão da variabilidade desta estimativa para os dois tipos de coleta de dados: **amostras emparelhadas e independentes**.

Coonsideremos uma situação em que os pacientes do estudo estão divididos em dois grupos e a resposta de interesse é contínua.

O comportamento global em cada grupo é sintetizado pela *média aritmética* dos valores da variável resposta.

O *efeito do tratamento* é definido como a diferença das médias das respostas dos pacientes nos dois grupos.

8.2.1 Amostras pareadas

A variável resposta é denotada por X_1 e X_2 para os dois grupos a serem comparados.

- Os dados são pares de observações: $(x_{11}, x_{12}), (x_{21}, x_{22}), \cdots, (x_{n1}, x_{n2})$.
- O efeito da intervenção para cada par é medido pelas diferenças $d_1 = x_{11} - x_{12}, d_2 = x_{21} - x_{22}, \cdots, d_n = x_{n1} - x_{n2}$;

- a *estimativa do efeito* $\mu_1 - \mu_2$ é dada por

$$\bar{d} = \bar{x}_1 - \bar{x}_2$$

- a variabilidade associada a este valor é dada pelo *intervalo de confiança* para $\mu_1 - \mu_2$, que neste caso é

$$\left[\bar{d} - t^* \frac{s_d}{\sqrt{n}}, \bar{d} + t^* \frac{s_d}{\sqrt{n}}\right]$$

em que $t^*_{1-\alpha/2}$ é o percentil de ordem $1-\alpha/2$ da distribuição t com $n-1$ graus de liberdade e s_d é o desvio-padrão das diferenças, isto é, $s_d = \sqrt{\frac{\sum(d_i - \bar{d})^2}{n-1}}$.

Exemplo: Avaliação de redução da pressão intraocular

Em estudo realizado no hospital São Geraldo da UFMG, Cronemberg & Calixto (1991) estudaram a capacidade de redução da pressão intraocular das drogas timolol, betaxolol e levobunolol.

Para isto, medicaram 10 pacientes com as três drogas e compararam os resultados com aqueles obtidos quando os pacientes receberam um placebo.

As medidas da pressão intraocular, expressas em $mmHg$, obtida às 6 horas da manhã com os pacientes usando timolol e placebo são apresentadas na Tabela 24.

Placebo	22	25	23	18	24	24	17	23	22	23
Timolol	18	20	20	17	16	20	20	20	20	24
Efeito	-4	-5	-3	-1	-8	-4	3	-3	-2	1

Tabela 24: Pressão ocular de pacientes em uso de placebo e timolol

- $\bar{d} = -2,6mmHg$ é uma estimativa do efeito hipotensor do timolol, ou seja, esses 10 pacientes fornecem indícios de que o timolol diminui a pressão ocular em 2,6 mmHg.
- $s_d = 3,098, s_d/\sqrt{n} = 0,98, t^* = 2,262$. O intervalo de confiança com nível de 95% para o efeito hipotensor do timolol é

$$[-2,6 - (2,262)(0,98); -2,6 + (2,262)(0,98)] = [-4,8; -0,4]$$

Podemos afirmar que o timolol reduz a pressão ocular por uma quantidade que varia em média de 0,4 mmHg a 4,8 mmHg.

- a não inclusão do zero no intervalo acima indica, ao nível de confiança de 95%, a existência de efeito. Este resultado é o mesmo que obteríamos se através do uso do teste t para dados pareados.

8.2.2 Amostras independentes

- Denotamos as observações do primeiro grupo de $x_{11}, x_{21}, \cdots, x_{n_1 1}$ e as do segundo $x_{21}, x_{22}, \cdots, x_{n_2 2}$;
- a *estimativa do efeito* $\mu_1 - \mu_2$ é dada por

$$\bar{x}_1 - \bar{x}_2$$

- a variabilidade associada a este valor é dada pelo *intervalo de confiança* para $\mu_1 - \mu_2$, que neste caso é obtido usando o seguinte resultado

$$\frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t_{n_1+n_2-2}$$

em que

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

Obtendo da tabela da distribuição t com $n_1 + n_2 - 2$ graus de liberdade um número t^* tal que $P(-t^* \leq t_{n_1+n_2-2} \leq t^*) = 1 - \alpha$ onde $1 - \alpha$ é o nível de confiança escolhido.

Podemos escrever o intervalo de confiança para $\mu_1 - \mu_2$ como

$$\left[(\bar{x}_1 - \bar{x}_2) - t^* s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}; (\bar{x}_1 - \bar{x}_2) + t^* s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right]$$

Exemplo: Comparação da tianeptina com placebo (continuação)

Podemos obter uma estimativa do efeito da tianeptina ao fim de 42 dias de uso.

Grupo placebo	$n_1 = 15$	$\bar{x}_1 = 20,53$	$s_1 = 11,09$
Grupo tianeptina	$n_2 = 16$	$\bar{x}_2 = 11,37$	$s_2 = 7,26$

1. Estimativa do efeito da droga: 20,53-11,37=9,16
2. Intervalo de 95% de confiança:

$$\left[9,16 \pm (2,0452)(9,31)\sqrt{\frac{1}{15} + \frac{1}{16}}\right] = [2,32; 16,00]$$

A tianeptina tem claro efeito antidepressivo, ainda que medido com grande variabilidade neste ensaio clínico.

8.3 Medida do efeito: resposta dicotômica

A variável resposta dicotômica é o tipo padrão em estudos caso-controle e de coorte.

A Tabela 25 apresenta as proporções dos pacientes correspondentes a cada possível classificação.

Doença	Fator de exposição		
	Presente	Ausente	Total
Presente	pP_1	qP_0	$pP_1 + qP_0$
Ausente	pQ_1	qQ_0	$pQ_1 + qQ_0$
Total	p	q	1

Tabela 25: Proporções em estudos caso-controle ou coorte

- p e $q = 1 - p$, em estudos de coorte, indicam o tamanho relativo das coortes do estudo, ou seja, $p = P(E)$ e $q = P(\bar{E})$.
- P_1 é a proporção dos pacientes que desenvolveram a doença entre os pacientes expostos, ou seja, $P_1 = P(D|E)$. Assim, $pP_1 = P(E)P(D|E) = P(D \cap E)$.
- P_0 é a proporção dos pacientes que desenvolveram a doença entre os pacientes não expostos, ou seja, $P_0 = P(D|\bar{E})$. Então, $qP_0 = P(\bar{E})P(D|\bar{E}) = P(D \cap \bar{E})$.

Para exemplificar, vamos supor que $P_1 = 0,20, P_0 = 0,10$ e $p = 0,4$.

Doença	Fator de exposição		
	Presente	Ausente	Total
Presente	0,08	0,06	0,14
Ausente	0,32	0,54	0,86
Total	0,4	0,6	1

Tabela 26: Valores da Tabela 25 com $P_1 = 0,20$, $P_0 = 0,10$ e $p = 0,4$

Os dados amostrais provenientes destes estudos são usualmente apresentados como na Tabela 27.

Doença	Fator		
	Presente	Ausente	Total
Presente	a	b	$a + b$
Ausente	c	d	$c + d$
Total	$a + c$	$b + d$	N

Tabela 27: Frequências em estudos caso-controle ou coorte

Os estimadores de P_1 e P_0 são: $\hat{P}_1 = a/(a + c)$ e $\hat{P}_0 = b/(b + d)$.

8.3.1 Risco relativo

Uma medida natural para quantificar o efeito da exposição ao fator é a razão conhecida como risco relativo denotado por RR é:

$$RR = P(D|E)/P(D|\bar{E}) = P_1/P_0$$

que, em estudos de coorte, pode ser estimado por

$$\hat{RR} = \hat{P}_1/\hat{P}_0 = \frac{\frac{a}{a+c}}{\frac{b}{b+d}}$$

Este mesmo estimador também pode ser utilizado para estudos em forma de ensaios clínicos pois estes podem ser vistos como estudos de coorte, em que os grupos são criados através de alocação aleatória de pacientes aos grupos.

Exemplo: Efeito preventivo da aspirina

- Objetivo do estudo: testar se 325 mg de aspirina (65% de um comprimido usual que é de 0,5 g), ingeridas em dias alternados, reduzem a mortalidade devida a doenças cardiovasculares.
- Coleta de dados:
  - ensaio clínico controlado, duplo-cego, tempo de seguimento médio de 57 meses,
  - 22071 médicos americanos: 11037 receberam aspirina ativa e 11034 o placebo,
  - com idades entre 40 e 84 anos,
  - não tinham história de infarto do miocárdio, AVC ou ataque isquêmico transitório,
  - não usavam regularmente aspirina e nem apresentavam contra-indicações ao seu uso.
- Resultados:foram observados 139 infartos no grupo que tomava aspirina e 239 no grupo placebo: $\hat{P}_1 = \frac{139}{11037} = 0,013$ e $\hat{P}_0 = \frac{239}{11034} = 0,022$. Uma estimativa do risco relativo é:

$$\hat{RR} = \frac{0,013}{0,022} = 0,59 \rightarrow \frac{1}{\hat{RR}} = \frac{0,022}{0,013} = 1,69$$

  O risco de quem tomava aspirina regularmente é 59% do risco dos que não o faziam, ou o risco de quem não toma a droga é 1,69 vezes maior do que o dos usuários.

8.3.2 Razão de odds

O risco relativo não pode ser estimado em estudos de caso-controle porque, neste tipo de estudos, as incidências observadas são meras consequências do número escolhido de casos e controles e não características dos grupos em estudo.

Para exemplificar, vamos supor que os dados resultantes de um estudo caso-controle (esquema 1:1): Utilizando o estimador de P_1 e P_0 obtemos: $\hat{P}_1 = \frac{80}{110} = 0,73$ e $\hat{P}_0 = \frac{20}{90} = 0,22$.

Grupo	Fator de exposição		
	Presente	Ausente	Total
Casos	80	20	100
Controles	30	70	100
Total	110	90	200

Tabela 28: Frequências observadas num estudo hipotético de caso-controle (esquema 1:1)

Grupo	Fator de exposição		
	Presente	Ausente	Total
Casos	80	20	100
Controles	60	140	200
Total	140	160	300

Tabela 29: Frequências observadas num estudo hipotético de caso-controle (esquema 1:2)

Se no entanto, tivessemos definidos dois controles para cada caso (esquema 1:2), teríamos a seguinte tabela:

Utilizando o estimador de P_1 e P_0 obtemos desta vez: $\hat{P}_1 = \frac{80}{140} = 0,57$ e $\hat{P}_0 = \frac{20}{160} = 0,13$.

Para este tipo de estudos define-se o efeito da exposição de um forma alternativa denominada **razão das odds** (Odds Ratio):

- Define-se a *odds* de se desenvolver a doença entre os expostos por $\frac{P_1}{Q_1}$ e entre os não expostos como $\frac{P_0}{Q_0}$.
- A *razão das odds* (ψ) é:
$$\psi = \frac{\frac{P_1}{Q_1}}{\frac{P_0}{Q_0}} = \frac{P_1 Q_0}{P_0 Q_1}$$
- O estimador de ψ é dado por
$$\hat{\psi} = \frac{\frac{a/(a+c)}{c/(a+c)}}{\frac{b/(b+d)}{d/(b+d)}} = \frac{ad}{bc}$$

Por razões teóricas a variação de $\hat{\psi}$ é mais facilmente calculada na escala logaritmica. Pode-se provar que $ln\hat{\psi}$ tem aproximadamente distribuição gaussiana com média $ln\psi$ e variância estimada por:
$$Var(ln\hat{\psi}) = \frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$$

Podemos assim construir intervalos de confiança para $ln\psi$
$$\left[ln\hat{\psi} - z^*\sqrt{Var(ln\hat{\psi})}; ln\hat{\psi} + z^*\sqrt{Var(ln\hat{\psi})}\right]$$

onde Z^* é o percentil obtido da distribuição gaussiana padrão tal que $P(-z^* < Z < z^*) = 1-\alpha$.

Se este intervalo contém o zero (correspodendo ao valor 1 para a razão de odds) então a associação entre a doença e o fator não é estatisticamente significativa.

O intervalo de confiança para ψ é obtido exponenciando-se os limites inferior e superior do intervalo.

Motivos para a adoção da razão de odds como a forma de se medir associação entre fator de risco e doença:

1. Usualmente as doenças são raras, isto é, P_1 e P_0 são pequenos e portanto $Q_1 = Q_0 \approx 1$. Nesta situação, $\psi \approx \frac{P_1}{P_0} = RR$.

2. ψ pode ser estimado com dados de qualquer tipo de estudo.

Exemplo: Amamentação na infância e câncer de mama

Para verificar se o fato de ter sido amamentado pela mãe é um fator de proteção para o câncer de mama, Freudenheim et al. (1994) realizaram estudo do tipo caso-controle nos condados de Erie e Niágara situados na parte oeste do estado de Nova York (EUA).

As pacientes tomadas como controle foram escolhidas na população da região, não havendo emparelhamento.

Os dados obtidos estão apresentados na Tabela 30.

Grupo	Amamentação	
	Sim	Não
Casos	353	175
Controles	449	153

Tabela 30: Distribuição de casos e controles segundo a amamentação

O risco de desenvolver câncer de mama entre mulheres amamentadas pela mãe, aproximado pela razão de odds, é estimado por

$$\hat{\psi} = \frac{(353)(153)}{(175)(449)} = 0,69$$

o risco do grupo amamentado é 69% do risco de grupo não amamentado.

Para obtermos o intervalo de confiança para ψ temos que calcular:

$$ln\hat{\psi} = ln(0,69) = -0,37$$

$$Var(ln\hat{\psi}) = \frac{1}{353} + \frac{1}{175} + \frac{1}{449} + \frac{1}{153} = 0,02$$

Um intervalo de confiança de 95% para $ln\psi$ é

$$\left[-0,37 - 1,96 \times \sqrt{0,02}; -0,37 + 1,96 \times \sqrt{0,02}\right] = [-0,64; -0,10]$$

e o intervalo de confiança para ψ é portanto

$$[exp(-0,64); exp(-0,10)] = [0,53; 0,90]$$

indicando uma associação significativa entre ter sido amamentada e câncer de mama.

Esse resultado deve ser interpretado com cuidado, uma vez que não foram considerados fatores importantes, como história familiar e idade na primeira gestação. De fato, ao ajustar o modelo incorporando essas variáveis a associação deixou de ser significativa.

9 Comparando dois grupos

Uma questão importante que surge no trabalho de pesquisa na área da saúde é a comparação de drogas, de métodos cirúrgicos, de condições experimentais, de procedimentos de laboratórios, de dietas ou, em geral, de tratamentos.

Um caso especial que ocorre frequentemente é o da comparação de **dois** tratamentos. O objetivo pode ser o de se estabelecer a superioridade de um tratamento ou a equivalência entre eles.

A escolha entre dois tratamentos é menos simples do que em princípio parece.

Isto porque os seres vivos geralmente reagem de forma diferente a um tratamento. O resultado de um tratamento pode variar enormemente de indivíduo para indivíduo. Como não se conhece a priori a reação de cada indivíduo, em geral, considera-se como tratamento mais eficiente aquele que *na média* fornece os melhores resultados.

Em outras palavras, a situação ideal da escolha do melhor tratamento para cada indivíduo não é possível na prática.

Consequentemente, considera-se como o melhor tratamento aquele que produz bons resultados para a grande maioria da população em estudo.

Extrapolações dos resultados obtidos no grupo estudado para uma população de interesse exige cuidados especiais no planejamento tanto de ensaios clínicos quanto de estudos observacionais. Fatores que afetam a resposta sendo avaliada devem ser controlados, tanto na fase de coleta como na análise dos dados. Por exemplo, no estudo da pressão arterial, a idade, a raça e o sexo são fatores que devem ser levados em conta.

O procedimento para determinar qual dos dois tratamentos é em média o mais eficiente envolve geralmente a seleção de duas amostras e a comparação dos resultados obtidos.

Vamos discutir como comparar os efeitos médios de dois tratamentos, quando sua resposta é medida para uma variável contínua ou dicotômica.

A seguir serão apresentados os testes mais comuns para quatro situações: *variável dicotômica* (amostras independentes e pareadas) e *variável contínua* (amostras independentes e pareadas).

9.1 Conceitos fundamentais

O procedimento estatístico denominado *Teste de Hipóteses* é usado amplamente nas áreas do conhecimento humano em que as variáveis envolvidas estão sujeitas a variabilidade. Embora as idéias básicas desta seção sejam de aplicação geral, serão ilustradas apenas no contexto da comparação de dois tratamentos médicos.

Exemplo: Eficácia do AZT

Fisch et al. (1987) publicaram o primeiro relato de um ensaio clínico que comprovou a eficácia de zidovudina (AZT) para prolongar a vida de pacientes com AIDS. Os dados centrais do trabalho estão na Tabela 31.

A análise do dados da Tabela 31 consiste basicamente na comparação de duas proporções.

	Situação		
Grupo	Vivo	Morto	Total
AZT	144	1	145
Placebo	121	16	137
Total	265	17	282

Tabela 31: Número de sobreviventes tratados com AZT ou placebo

A proporção dos que estavam vivos depois de 24 semanas de tratamento foi de 144/145=0,993 entre os pacientes que receberam o AZT, enquanto que para o grupo placebo foi de 121/137=0,883.

Como a alocação dos pacientes aos grupos foi feita de forma aleatória, a diferença entre essas duas proporções parece indicar que em pacientes com AIDS o AZT tem o efeito de prolongar a vida.

Antes de aceitar esta conclusão, entretanto, é preciso **afastar o acaso como explicação alternativa.** Ou seja, deve-se responder à pergunta:

Será que este resultado ocorreu por mero acaso ou por ser o AZT de fato uma droga efetiva?

9.1.1 Hipóteses a serem testadas

Geralmente podemos formular os problemas, como o do AZT, através das hipóteses seguintes:

- **Hipótese nula**

  No problema de comparação de dois tratamentos é usual fixar como hipótese de interesse a inexistência de diferença entre os dois tratamentos comparados.

  Como frequentemente a comparação é feita entre um tratamento padrão e um tratamento novo, esta opção implica colocar o ônus da prova de efetividade no tratamento novo, uma opção conservadora mas prudente.

  Por esta razão, a hipótese a ser testada é usualmente chamada hipótese nula (H_0), nome que se generalizou mesmo para situações em que o problema não é mais de comparação entre dois tratamentos.

- **Hipótese alternativa**

  A hipótese nula deve ser comparada com uma hipótese alternativa, denominada H_1. Para cada situação existem muitas hipóteses alternativas adequadas. Entretanto, seguindo convenção estabelecida pelos editores de revistas científicas na área médica, a hipótese alternativa será sempre a inexistência de igualdade entre os tratamentos.

  No exemplo do AZT a hipótese nula é

  $$H_0 : p_C = p_T$$

  em que p_C e p_T são respectivamente as probabilidades de se observar a resposta de interesse entre os controles e entre os pacientes do grupo tratamento.

  De acordo com a opção feita acima, a hipótese alternativa será

$$H_1 : p_C \neq p_T$$

9.1.2 Critério de decisão

Decididas as hipóteses as serem testadas, o próximo passo é construir um critério baseado no qual a hipótese H_0 será julgada.

O critério de decisão é baseado na *estatística de teste*.

De uma forma bem genérica e intuitiva podemos dizer que a *estatística de teste mede a discrepância entre o que foi observado na amostra e o que seria esperado se a hipótese nula fosse verdadeira*.

Uma grande distância medida pela distribuição de probabilidade é indicação de que H_0 não é verdadeira, devendo portanto ser rejeitada.

Resumindo, rejeita-se a hipótese nula se o valor da estatística de teste é "grande". Esse valor deve portanto ser comparado a alguma distribuição de probabilidade.

9.1.3 Erros do tipo I e II e nível de significância

No exemplo do AZT havia a possibilidade de se rejeitar a hipótese de igualdade entre o AZT e o placebo, mesmo se de fato eles fossem iguais.

- *Erro do tipo I*: decisão de rejeitar H_0 quando de fato H_0 é verdadeira.

  Para evitá-lo, escolhemos umcritério de decisão que torna este erro pouco provável.

- *Nível de significância do teste*: probabilidade de cometer o erro do tipo I, usualmente representado pela letra grega α.

Há, no entanto, um *segundo tipo de erro*. No exemplo do AZT ele consiste em não rejeitar a hipótese de igualdade entre o AZT e o placebo quando de fato estes dois tratamentos são diferentes. Isto implicaria na não liberação do novo tratamento, cujo efeito real não estaria sendo percebido.

- *Erro do tipo II*: decisão de não rejeitar H_0 quando de fato H_0 é falsa.

Para um tamanho fixo da amostra, não há como controlar simulataneamente ambos os erros. Convencionou-se que o erro mais sério seria o erro do tipo I.

Em um segundo momento, calcula-se o tamanho da amostra que reduza a probabilidade do erro do tipo II, usualmente representado pela letra grega β.

A Tabela 32 sintetiza os erros possíveis

A capacidade de um teste identificar diferenças que realmente existem, ou seja, de rejeitar H_0 quando é realmente falsa, é denominada *poder do teste* e é definida como 1-β.

Situação	Conclusão do teste	
real	Rejeitar H_0	Não rejeitar H_0
H_0 Verdadeira	erro tipo I	decisão correta
H_0 Falsa	decisão correta	erro tipo II

Tabela 32: Erros possíveis associados a teste de hipóteses

9.1.4 Probabilidade de significância (valor-p)

Existem duas opções para expressar a conclusão final de um teste de hipóteses.

1. Comparar o valor da estatística do teste com o valor obtido a partir da distribuição teórica, específica para o teste, para um valor pré-fixado do nível de significância;
2. Quantificar a chance do que foi observado ou resultados mais extremos, sob a hipótese de igualdade dos grupos. Essa opção baseia-se na probabilidade de ocorrência de valores iguais ou superiores ao assumido pela estatística do teste, sob a hipótese de que H_0 é verdadeira. Este número é chamado de *probabilidade de significância* ou *valor-p* e frequentemente é indicado apenas por p.

Como o valor-p é calculado supondo-se que H_0 é verdadeira, pode-se fazer duas conjecturas quando se obtem um valor muito pequeno.

- Um evento que é extremamente raro pode ter ocorrido; ou
- a hipótese H_0 não deve ser verdadeira, isto é, a conjectura inicial e conservadora não parece plausível.

Portanto, quanto menor o valor-p maior a evidência para se rejeitar H_0.

De um modo geral, na área médica, considera-se que valor-p menor ou igual a 0,05 indica que há diferenças significativas entre os grupos comparados.

9.2 Resposta dicotômica: amostras independentes

Compara dois grupos através do resultado observado em uma variável dicotômica é um problema comum na pesquisa médica, aparecendo com frequência em todos os tipos de estudos clínicos.

A variável de interesse é a ocorrência de um evento, como o desenvolvimento de uma doença, ou a presença de certo atributo, por exemplo, albinismo.

O problema da comparação das probabilidades de ocorrência do evento ou do atributonos dois grupos (p_1 e p_2) é formulado através das hipóteses

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2$$

9.2.1 Teste qui-quadrado (χ^2)

A Tabela 33 apresenta dados genéricos de uma situação envolvendo a comparação de dois grupos e que a resposta de interesse é dicotômica: a ocorrência ou não de um evento.

Grupo	Ocorrência do evento		
	Sim	Não	Total
I	a	b	$a+b=n_1$
II	c	d	$c+d=n_2$
Total	$m_1=a+c$	$m_2=b+d$	$n_1+n_2=N$

Tabela 33: Distribuição quanto à ocorrência de um evento

Se não há diferença entre as proporções de ocorrência do evento nos dois grupos, então

$$\frac{a}{n_1}=\frac{c}{n_2}=\frac{a+c}{n_1+n_2}=\frac{m_1}{N}$$

A partir dessas igualdades podemos escrever

$$a=\frac{m_1\times n_1}{N}, b=\frac{m_2\times n_1}{N}, c=\frac{m_1\times n_2}{N}, d=\frac{m_2\times n_2}{N}$$

Temos, portanto, dois conjunto de valores: os **observados** (O_i) e os **esperados** (E_i) calculados sob a hipótese de igualdade das proporções de sucesso entre os grupos.

Se as proporções de sobrevivência são iguais nos dois grupos, a discrepância entre os dois conjuntos de números acima não deve ser grande.

Pearson propôs medir a discrepância entre os valores observados e esperados das quatro entradas de uma tabela 2×2 através da expressão

$$X^2=\sum_{i=1}^{4}\frac{(O_i-E_i)^2}{E_i}$$

É possível mostrar que o valor do X^2 pode ser calculado de maneira fácil e algebricamente equivalente através de

$$X^2=\frac{N(ad-bc)^2}{m_1m_2n_1n_2}$$

Exemplo: Eficácia do AZT (continuação)

Supondo que o AZT e o placebo são equivalentes, teríamos os valores esperados mostrados na Tabela 35.

	Situação		
Grupo	Vivo	Morto	Total
AZT	144	1	145
Placebo	121	16	137
Total	265	17	282

Tabela 34: Número de sobreviventes tratados com AZT ou placebo

	Situação		
Grupo	Vivo	Morto	Total
AZT	136,26	8,74	145
Placebo	128,74	8,26	137
Total	265	17	282

Tabela 35: Número esperado de sobreviventes sob a hipótese de equivalência entre AZT e placebo

Os cálculos necessários para a obtenção do valor da discrepância proposta por Pearson são mostrados na Tabela 36.

i	O_i	E_i	$O_i - E_i$	$(O_i - E_i)^2$	$\frac{(O_i - E_i)^2}{E_i}$
1	144	136,26	7,74	59,91	0,44
2	121	128,74	-7,74	59,91	0,47
3	1	8,74	-7,74	59,91	6,85
4	16	8,26	7,74	59,91	7,25
Total	282	282	0	239,64	15,01

Tabela 36: Cálculos necessários para a construção do teste χ^2

Portanto, $X^2 = 0,44 + 0,47 + 6,85 + 7,25 = 15,01$.

Calculado o valor de X^2, é preciso decidir se este é ou não um valor "grande".

Para se tomar uma decisão sobre a igualdade ou não das proporções, é preciso conhecer a distribuição de X^2 sob a hipótese nula. Esta distribuição foi obtida e recebeu o nome de **qui-quadrado** com 1 grau de liberdade, é indicada por χ^2 e está tabelada.

O valor da estatística de teste foi de 15,01. Como este valor é maior do que 3,84, valor obtido da distribuição do χ^2 para um nível de significância de 5%, rejeitamosa hipótese de igualdade entre os grupos de tratamento e controle. Em outras palavras, decidimos com 95% de certeza que há evidência de efeito do AZT.

Para obtermos o valor-p devemos calcular a probabilidade de encontrar valores maiores que 15,01, isto é, $P(\chi^2 \geq 15,01)$, sendo verdadeira a hipótese de igualdade das proporções.

Da tabela da distribuição χ^2, vemos que este valor é aproximadamente 0,0001.

Baseado neste estudo podemos dizer com grande certeza que o AZT tem efeito de prolongar a vida de pacientes com AIDS, primeira evidência necessária para a liberação do medicamento.

Exemplo: Fatores de risco para AVC

Wellin et al. (1987) relataram os resultados de um estudo realizado como objetivo de determinar os fatores de risco para acidente vascular cerebral (AVC) em homens de meia idade.

A Tabela 37 apresenta as proporções de presença de alguns fatores de risco nos indivíduos que sofreram AVC e naqueles que não foram acometidos pela doença.

Fator	AVC		
	sim	não	Valor-p
AVC como causa principal da morte			
na mãe	29,8	11,2	0,0005
no pai	7,0	7,5	0,73
AVC como causa principal ou contribuindo para a causa da morte			
na mãe	29,8	14,2	0,002
no pai	7,0	9,2	0,59
Doença coronariana	7,0	6,3	0,83
Sinais eletrocardiográficos			
hipertrofia do ventrículo esquerdo	3,5	1,0	0,08
doença coronariana	10,5	6,5	0,23

Tabela 37: Percentuais e valor-p para a comparação do grupo que sofreu com o que não sofreu AVC

A ocorrência de AVC na mãe como causa principal de morte pode ser considerada um possível fator de risco para o aparecimento de AVC no filho, uma vez que a diferença observada foi altamente significativa (p=0,0005).

9.2.2 Teste exato de Fisher

Há uma dificuldade técnica na aplicação do teste qui-quadrado quando o valor esperado em alguma casela na tabela 2×2 é menor que 5. Neste caso, o uso da distribuição χ^2_1 não é mais completamente apropriado. Ou seja, o grau de certeza na decisão tomada não é exatamente aquele fornecido pela distribuição χ^2_1.

A alternativa é usar o teste exato de Fisher (disponível na maioria dos programas de análise estatística), que é a versão exata do teste qui-quadrado. Não discutiremos aqui as etapas de sua construção, já que estas são muito técnicas para o nível pretendido neste curso.

Apresentamos o resultado do teste aplicado a uma situação em que seu uso é apropriado.

Exemplo: Tratamento de pneumonia bacteriana

Pasternak et al. (1992) avaliaram a eficácia e segurança de dois antibióticos no tratamento de pneumonia bacteriana de origem comunitária em adultos. Foram estudados 63 pacientes, sendo 32 tratados com cefadroxil e 31 com cefalexina.

A avaliação da resposta terapêutica foi baseada na evolução do quadro clínico e do exame radiológico do tórax feito na admissão ao estudo e no 10^o dia de tratamento.

Dos pacientes avaliados, a cura completa ocorreu em 31 dos 32 (96,9%) pacientes do grupo cefadroxil e em 28 dos 31 (90,3%) pacientes do grupo cefalexina.

Foram observados efeitos adversos em quatro casos: um (3,1%) no grupo que recebeu cefadroxil, com ocorrência de náuseas que desapareceram com o tratamento sintomático instituído; todos os três casos (9,7%) do grupo cefalexina tiveram diarréia, dois se recuperaram espontaneamente e um teve que interromper o tratamento.

Para comparar os percentuais de cura e de ocorrência de efeitos adversos, por se tratar de amostras pequenas, utilizou-se o teste exato de Fisher.

Quadro	Cefadroxil	Cefalexina	valor-p
Cura	96,9%	90,3%	0,5886
Efeitos adversos	3,1%	9,7%	0,5898

Tabela 38: Percentuais de curas, de efeitos colaterais e valor-p na comparação de cefadroxil e cefalexina

Não há evidência de diferença entre os dois tratamentos tanto em termos de cura como na ocorrência de efeitos adversos.

9.3 Resposta dicotômica: amostras pareadas

Suponhamos que dois patologistas examinaram, separadamente, o material de 100 tumores e os classificaram como benignos ou malignos.

Questão de interesse: os patologistas diferem nos seus critérios de decisão?

Diagnóstico	Diagnóstico de A		
de B	Malignos	Benignos	Total
Malignos	9	1	10
Benignos	9	81	90
Total	18	82	100

Tabela 39: Classificação de dois patologistas (A e B) quanto à malignidade de tumores

1. A unidade de análise é o tumor, avaliado por dois patologistas.
2. Foram feitas 200 análises, mas o total de tumores é 100.
3. Alguns tumores são mais malignos do que outros e, portanto, a hipótese fundamental na construção do teste de probabilidade constante de malignidade não é razoável aqui.

9.3.1 Teste de McNemar

Este teste aparece com grande frequência na análise de dados pareados resumidos no formato da Tabela 40.

Controle	Tratamento		
	Sucesso	Fracasso	Total
Sucesso	k	r	n_1
Fracasso	s	l	n_2
Total	m_1	m_2	N

Tabela 40: Resultados de uma classificação de dados pareados

A nomenclatura adotada é sucesso e fracasso para a ocorrência ou não do evento de interesse.

Se p_1 e p_2 são as probabilidades de sucesso nos grupos controle e tratamento, respectivamente, a hipótese de interesse é

$$H_0 : p_1 = p_2 \text{ versus } H_1 : p_1 \neq p_2$$

Os pares que produziram sucesso ou fracasso nos dois elementos do par não contém informação para discriminar p_1 de p_2, mas sim os pares de resultados discordantes (r e s).

Se H_0 for verdadeira espera-se que as discordâncias observadas sejam fruto do caso. Em outras palavras, sob H_0 espera-se a metade do número de discordâncias $\frac{r+s}{2}$.

A hipótese H_0 deve, portanto, ser rejeitada se a distância entre os valores discordantes observados e os esperados for grande.

A estatística do teste, usando uma correção de continuidade, é

$$X^2_{McN} = \frac{\left(|r - \frac{r+s}{2}| - \frac{1}{2}\right)^2}{\frac{r+s}{2}} + \frac{\left(|s - \frac{r+s}{2}| - \frac{1}{2}\right)^2}{\frac{r+s}{2}} = \frac{\left(|r-s| - 1\right)^2}{r+s}$$

O teste consiste em se rejeitar H_0 quando

$$X^2_{McN} > \chi^2_{1,1-\alpha}$$

em que $\chi^2_{1,1-\alpha}$ é o percentil de ordem $1-\alpha$ da distribuição qui-quadrado com 1 grau de liberdade.

Exemplo: Quimioterapia e câncer de mama

Dois tratamentos (A e B) para câncer de mama foram comparados. O grupo A recebeu tratamento quimioterápico durante seis meses começando na primeira semana após a mastectomia, enquanto que o grupo B, só na primeira semana.

Para que os grupos sejam tão comparáveis quanto possível com relação a fatores prognósticos, foram formados pares de pacientes com aproximadamente mesma idade e condições clínicas.

Os dados da Tabela 41 mostram que as proporções de sobrevivência após 5 anos de pacientes dos dois grupos são bastante parecidas, mas estes cálculos não levam em consideração o pareamento.

Tratamento	Sobrevivência		Total	% de
	Sim	Não		sobreviventes
A	526	95	621	$\hat{p}_A = 0,847$
B	515	106	621	$\hat{p}_B = 0,829$
Total	1041	201	1242	$\hat{p}_{A+B} = 0,816$

Tabela 41: Distribuição de sobrevivência segundo tratamento

Somos tentados a aplicar o teste qui-quadrado de Pearson nessa tabela ($X^2 = 0,72$), mas não é apropriado nessa situação em que obviamente as amostras **não são independentes**.

A Tabela 42 mostra os dados na forma de pares.

Sobreviência do	Sobrevivência do tratamento B		
tratamento A	Sim	Não	Total
Sim	510	16	526
Não	5	90	95
Total	515	106	621

Tabela 42: Distribuição de sobrevivência dos pares

Como $\chi^2_{McN} = \frac{(|16-5|-1)^2}{16+5} = 4,76\,(p = 0,029)$, há evidência estatística de diferença entre os tratamentos. Pode-se mostrar que o tratamento A fornece melhores resultados.

9.4 Resposta contínua: amostras independentes

Apresentamos aqui a metodologia para comparar dois grupos de pacientes em relação a uma resposta contínua, por exemplo, pressão sistólica.

Testa-se, neste caso, a igualdade das médias das respostas de dois tratamentos.

Sejam μ_1 e μ_2 as médias da variável estudada para os dois grupos. As hipóteses a serem testadas são

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2$$

9.4.1 Teste t

O teste t para duas amostras é adequado para situações em que as respostas aos dois tratamentos são variáveis quantitativas com distribuição gaussiana com parâmetros (μ_1, σ) e (μ_2, σ).

Suposição do teste: as variáves estudadas têm distribuições gaussianas com o mesmo desvio-padrão.

Note que μ_1, μ_2 e σ são parâmetros populacionais e, portanto constantes desconhecidas.

Para testar a hipótese acima,

1. coletamos uma amostra de tamanho n_1 no grupo 1 e uma amostra de tamanho n_2 no grupo 2.
2. A partir desses dados, calculamos as médias ($\bar{x}_1$ e $\bar{x}_2$) e os desvio-padrão (s_1 e s_2) dos dois grupos.
3. O critério de decisão para se testar a hipótese nula acima consiste em rejeitar H_0 se

$$T = \frac{\bar{X}_1 - \bar{X}_2}{DP(\bar{X}_1 - \bar{X}_2)}$$

é "grande", em que $DP(\bar{X}_1 - \bar{X}_2)$ é o desvio-padrão da diferença entre $\bar{X}_1$ e $\bar{X}_2$.

A decisão baseada nesta estatística de teste é intuitiva, pois

- quanto maior for a **diferença entre as médias** amostrais, maior a chance de estarmos diante de dois grupos realmente diferentes (captado pelo numerador de T).
- Por outro lado, quanto maior for a **variabilidade das respostas**, maior será a dificuldade de se detectar diferenças entre os efeitos médios (captado pelo denominador de T).

Em outras palavras, com a estatística T estamos medindo a diferença entre as médias em termos de desvios-padrão (lembra da interpretação do escore padronizado? É semelhante...).

Devemos rejeitar H_0 se T for "grande"em valor absoluto.

Para tomarmos esta decisão, usamos a distribuição t de Student (Student, 1908):

- Essa distribuição é caracterizada por um parâmetro que assume apenas valores inteiros positivos que são chamados graus de liberdade.
- É muito parecida com a distribuição gaussiana e sua forma típica é mostrada na Figura 44.
- Os percentis da distribuição t de Student para vários graus de liberdade são disponíveis em tabelas.
- Para graus de liberdade muito grandes, seus percentis são praticamente iguais aos da distribuição gaussiana.
- Na comparação de médias de amostras independentes, os graus de liberdade são n_1+n_2-2, o número total da amostra menos 2.

Para efetuarmos os cálculos necessários ao teste, precisamos conhecer a expressão para o desvio-padrão de $\bar{X}_1 - \bar{X}_2$.

$$DP(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma^2}{n_1} + \frac{\sigma^2}{n_2}} = \sigma \times \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

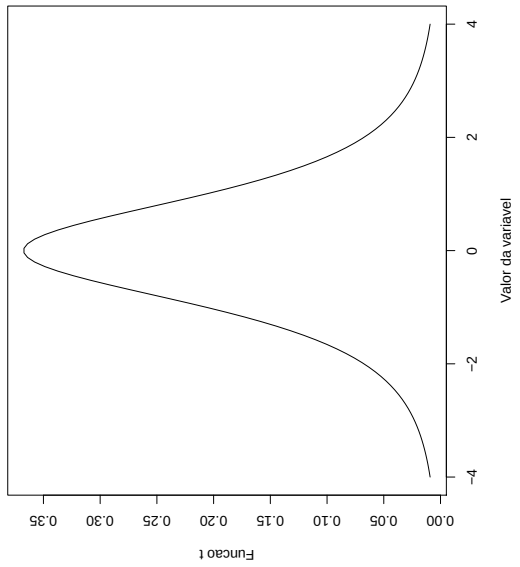


Figura 44: Distribuição t de Student

Precisamos encontrar uma estimativa de σ para obter uma estimativa do desvio-padrão de $(\bar{X}_1 - \bar{X}_2)$.

Como a suposição é que as variâncias dos dois grupos são iguais, podemos estimar σ como uma média ponderada, com pesos proporcionais aos tamanhos dos grupos,

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

O teste t para comparação de duas amostras consiste em se rejeitar H_0 em favor de H_1 ao nível de α de significância, se

$$\frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}| > t_{n_1+n_2-2;1-\frac{\alpha}{2}}$$

em que $t_{n_1+n_2-2;1-\frac{\alpha}{2}}$ é o percentil de ordem $1 - \alpha/2$ da distribuição t com $n_1 + n_2 - 2$ graus de liberdade.

Alternativamente, pode-se obter o valor-p da estatística de teste resultante dos dados e tomar a decisão com base neste p-valor.

Exemplo: Níveis séricos de frutosamina

A frutosamina é um índice do controle metabólico no diabetes mellitus podendo refletir as variações da glicemia nas últimas duas a três semanas. Representa um conjunto de proteínas glicosadas, cuja fração principal é a albumina.

Com o objetivo de estabelecer os valores normais da frutosamina em homens, mulheres gestantes ou não, Camargo et al.(1994) realizaram um estudo que incluia um grupo composto de 42

indivíduos normais, 21 mulheres (22 a 71 anos) e 21 homens (24 a 81 anos).

O outro grupo era constituído de 36 gestantes (17 a 37 anos) atendidas no ambulatório de Ginecologia e Obstetrícia do Hospital de Clínicas de Porto Alegre. A idade gestacional variou de 17 a 37 semanas.

Como recomendado, os autores apresentam as médias e desvios-padrão de todas as variáveis consideradas para cada grupo (Tabela 43 e 44), de forma que é possível repetir os cálculos de cada valor-p apresentado.

Variável	Mulheres ($n = 21$)		Homens ($n = 21$)	
	$\bar{x}_1$	s_1	$\bar{x}_2$	s_2
Idade (anos)	39	14	43	14
Glicose (mg/dL)	95	20	84	24
Frutosamina ($mmol/L$)	2,70	0,26	2,99	0,32
Triglicerídios (mg/dL)	145	135	124	62

Tabela 43: Média e desvio-padrão de variáveis para mulheres não-gestantes e homens

Variável	$\bar{x}$	s
Idade (anos)	25	7
Idade gestacional (semanas)	28	5
Glicose (mg/dL)	103	17
Frutosamina ($mmol/L$)	2,40	0,22
Albumina (mg/dL)	4,1	0,6

Tabela 44: Média e desvio-padrão de variáveis para gestantes

Resultados obtidos na comparação de homens e mulheres, através de teste t de amostras independentes:

- os grupos não diferem nas médias de idade ($p = 0,30$);
- a glicemia de jejum e a de 2 horas após a sobrecarga oral de glicose não foram diferentes ($p = 0,08$ e $p = 0,40$);
- diferença significativa nos níveis médios de frutosamina ($p = 0,002$).

Como a distribuição é gaussiana, a faixa de referência para frutosamina foi calculada como média $\pm$ 2 desvios-padrão, sendo portanto:

- (2,18 a 3,22) $mmol/L$ para mulheres;
- (2,35 a 3,63) $mmol/L$ para homens;
- (1,96 a 2,84) $mmol/L$ para gestantes.

Os achados principais indicam que devem ser considerados o sexo e a presença ou não de gravidez para se definir os limites de referência dos valores da frutosamina sérica.

Exemplo: Comparação da tianeptina com placebo

A tianeptina é um fármaco antidepressivo do grupo dos tricíclicos. Sua ação antidepressiva potencial foi demonstrada em estudos pré-clínicos através de testes em animais.

Rocha (1995) relata os resultados de um ensaio clínico aleatorizado, duplo-cego, realizado com o objetivo de comparar a tianeptina com o placebo.

Participaram deste ensaio pacientes de Belo Horizonte, Rio de Janeiro e Campinas.

O ensaio consistiu em administrar a droga a dois grupos de pacientes, compostos de forma aleatória, e quantificar a depressão através da escala de Montgomery-Asberg (MADRS) – os valores maiores indicam maior gravidade da depressão.

O escore foi obtido para cada paciente 7, 14, 21, 28 e 42 dias após o início do estudo.

A Tabela 45 apresenta os escores finais dos pacientes dos dois grupos admitidos em Belo Horizonte.

Grupo	Escores							
Placebo	6	33	21	26	10	29	33	29
	37	15	2	21	7	26	13	
Tianeptina	10	8	17	4	17	14	9	4
	21	3	7	10	29	13	14	2

Tabela 45: Escore final na escala MADRS de pacientes dos dois grupos admitidos em Belo Horizonte

Para se efetuar o teste t é preciso usar as seguintes informações:

$n_1 = 15$	$\bar{x}_1 = 20,53$	$s_1 = 11,09$
$n_2 = 16$	$\bar{x}_2 = 11,37$	$s_2 = 7,26$

Com estes valores temos que a estimativa de σ é dada por:

$$s_p = \sqrt{\frac{14(11,09)^2 + 15(7,26)^2}{15 + 16 - 2}} = 9,31$$

Finalmente,

$$T = |\frac{\bar{x}_1 - \bar{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}| = |\frac{20,53 - 11,37}{9,31\sqrt{\frac{1}{15} + \frac{1}{16}}}| = 2,74$$

que comparado com o valor de $t_{29;0,975} = 2,045$ leva à rejeição da igualdade entre os dois grupos no nível de 5%. O valor-p é 0,0104.

Exemplo: Fatores de risco para AVC (continuação)

No estudo de Gotemburgo, também foram consideradas diversas variáveis quantitativas como possíveis fatores de risco para AVC:

- fumo;
- pressão sistólica e diastólica;
- peso;
- altura;
- índice de massa corporal;
- medida de cintura;
- medida do quadril;
- índice de obesidade abdominal;
- nível sérico de colesterol;
- glicose no sangue;
- fibrinogênio no plasma;
- níveis de hemoglobina e de hematócrito;
- capacidade vital.

Na Tabela 46 estão reproduzidos os valores médios destas variáveis, divididos em dois grupos: os que sofreram AVC e os que não sofreram AVC.

Fator	AVC		
	Sim	Não	valor-p
Fumo	2,9	2,6	0,16
Pressão sistólica	151,9	143,2	0,004
Pressão diastólica	96,8	90,1	0,0001
Índice de massa corporal	25,7	24,9	0,08
Medida do quadril	94,8	93,6	0,23
Índice obesidade abdominal	0,95	0,93	0,0004
Colesterol	275	269	0,40
Glicose	64,8	67,9	0,60
Fibrinogênio	3,6	3,3	0,01
Hemoglobina	14,9	14,8	0,40
Hemtócrito	44,3	44,0	0,46
Capacidade Vital	4,3	4,5	0,03

Tabela 46: Valores médios dos fatores estudados no início da pesquisa e valor-p obtido através do teste t

Podemos dizer que a pressão sistólica, a pressão diastólica, o índice de obesidade abdominal, fibrinogênio e capacidade vital são possíveis fatores de risco para o AVC em homens.

A verificação final da influência destes fatores na ocorrência de AVC foi feita usando-se análise múltipla.

9.4.2 Teste Z para comparação de médias

O teste t que acabamos de descrever é baseado no pressuposto importante de que os dois grupos têm o mesmo desvio-padrão. Isso nem sempre acontece na prática.

No caso de amostras grandes (n_1 e $n_2 \geq 30$) dispomos de um teste em que não é necessário qualquer suposição adicional sobre σ_1 e σ_2, ou seja, os desvios podem ser iguais ou diferentes.

Para amostras grandes, $PD(\bar{X}_1 - \bar{X}_2)$ pode ser estimada por $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$.

A estatística do teste para testar a hipótese de igualdade de médias ($H_0 : \mu_1 = \mu_2$) é

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{s_1^2/n_1 + s_2^2/n_2}}$$

que é aproximadamente $N(0, 1)$, sob H_0, isto é, a hipótese nula é rejeitada se o valor absoluto da estatística for maior que $z_{1-\alpha/2}$, o percentil de ordem $(1 - \alpha/2)$ da distribuição $N(0, 1)$.

Exemplo: Efeito do halotano em cirurgias cardíacas

O halotano é uma droga bastante utilizada para induzir anestesia geral. Trata-se de um poderoso anestésico de inalação, não inflamável e não explosivo, com um odor relativamente agradável.

Pode ser administrado ao paciente com o mesmo equipamento usado para sua oxigenação.

Após a inalação,a substância chega aos pulmões tornando possível a passagem para o estado anestésico mais rapidamente do que seria possível com drogas administradas de forma intravenosa.

Entretanto, os efeitos colaterais incluem a depressão do sistema respiratório e cardiovascular, sensibilização a arritmias produzidas por adrenalina e eventualmente o desenvolvimento de lesão hepática.

Alguns anestesistas acreditam que esses efeitos podem causar complicações em pacientes com problemas cardíacos e sugerem o uso da morfina como um agente anestésico nesses pacientes devido ao seu pequeno efeito na atividade cardíaca.

Conahan et al.(1973) compararam esses dois agentes anestésicos em um grande número de pacientes submetidos a uma cirugia de rotin para reparo e substituição da válvula cardíaca. Para obter duas amostras comparáveis, os pacientes foram alocados aleatoriamente a cada tipo de anestesia.

A fim estudar o efeito desses dois tipos de anestesia, foram registradas variáveis hemodinâmicas, como pressão sanguínea antes da indução da anestesia, após a anestesia mas antes da incisão, e em outros períodos importantes durante a operação.

A questão que surge é se o efeito do halotano e da morfina na pressão sanguínea é o mesmo.

Para comparar os grupos necessitamos dos resultados apresentados na Tabela 47.

Devemos testar a diferença entre as pressões sanguíneas médias de indivíduos anestesiados com halotano ou morfina:

$$H_0 : \mu_1 = \mu_2 \text{ versus } H_1 : \mu_1 \neq \mu_2$$

Informações	Anestesia	
sobre a amostra	Halotano	Morfina
Média	66,9	73,2
Desvio-padrão	12,2	14,4
Tamanho (n)	61	61

Tabela 47: Média e desvio-padrão da pressão sanguínea segundo o tipo de anestesia

Como as amostras são grandes, podemos usar o teste Z

$$Z = |\frac{66,9 - 73,2}{\sqrt{12,2^2/61 + 14,4^2/61}}| = |\frac{-6,30}{\sqrt{5,84}}| = 2,61$$

Adotando um nível de significância de 5%, o resultado é estatísticamente significativo, já que $2,61 > 1,96$ (valor-p=0,009), indicando que os dois anestésicos não são equivalentes.

9.5 Resposta contínua: amostras pareadas

Consideremos a seguinte pergunta: *será que um tratamento baseado em uma dieta combinada com um programa de exercícios físicos é eficaz na redução no nível médio de colesterol total?*

Neste caso, o problema de interesse é uma comparação entre dois grupos de medidas de taxas de colesterol total.

Assumindo que a distribuição do nível de colesterol é gaussiana, e que o possível efeito do tratamento se dá apenas no aumento ou decréscimo na média da distribuição, não na sua variabilidade, então as hipóteses a serem testadas são

$$H_0 : \mu_A = \mu_D \text{ versus } H_1 : \mu_A \neq \mu_D$$

em que μ_A e μ_D são as médias antes e depois do tratamento.

- O **erro do tipo I**, é dizer que o tratamento tem efeito quando na realidade não tem.
- O **erro do tipo II**, é dizer que não há efeito quando na realidade tem.
- O **critério de decisão** utilizado para testar H_0 deve ser baseado na diferenças entre os valores do colesterol nas duas ocasiões. Se houver efeito do tratamento, então essas diferenças devem ser diferentes de zero.
- O problema da escolha de um critério de decisão reduz-se a escolher uma forma de verificar se as diferenças são provenientes de uma **distribuição com média zero**.

De uma maneira geral, considerando o par formado pelas medidas do indivíduo antes e depois do tratamento:

- os dados são n pares de observações;
- para cada par tomamos a diferença das duas observações (d);
- a partir dessas diferenças calculamos a média das diferenças

$$\bar{d} = \frac{\sum_{i=1}^{n} d_i}{n}$$

- e o desvio-padrão amostral das diferenças

$$s_d = \sqrt{\frac{\sum_{i=1}^{n} (d_i - \bar{d})^2}{n-1}}$$

- Um critério é a distância entre a média das diferenças $\bar{d}$ e zero.

$$T_p = \frac{\bar{D} - 0}{DP(\bar{D})}$$

em que $DP(\bar{D})$ é desvio-padrão de $\bar{D}$ com

$$DP(\bar{D}) = \frac{\sigma_d}{\sqrt{n}}$$

e σ_d é o desvio-padrão populacional das diferenças que é estimado por s_d, o desvio-padrão das diferenças na amostra.

- Portanto, o teste consiste em rejeitar H_0 se

$$T_p = \frac{\bar{d}}{s_d/\sqrt{n}}$$

for "grande", ou seja, se a distância entre a médias das diferenças e zero, medida em desvios-padrão tem um valor "grande".

A decisão de quão "grande"o valor de T_p deve ser exige o conhecimento de sua distribuição. Quando a distribuição de D é gaussiana, $T_p \sim t_{n-1}$.

- A regra do teste é então rejeitar H_0 se

$$|T_p| \geq t_{n-1,1-\alpha/2}$$

em que $t_{n-1,1-\alpha/2}$ é o percentil de ordem $(1-\alpha/2)$ da distribuição t de Student com $n-1$ graus de liberdade.

Note que:

- Quanto maior o valor $\bar{d}$ maior a evidência de que o tratamento reduz o nível de colesterol;
- Quanto menor a variabilidade das diferenças individuais maior a chance de se detectar um efeito médio significativo.

Exemplo: Programa para redução do nível de colesterol

Estudo com o objetivo de avaliar a efetividade de uma dieta combinada com um programa de exercícios físicos na redução do nível de colesterol.

Sejam μ_A e μ_D as médias dos níveis de colesterol antes e depois do programa.

Para testar a hipótese de que o programa altera o nível de colesterol

$$H_0 : \mu_A = \mu_D \text{ versus } H_1 : \mu_A \neq \mu_D$$

será aplicado o teste t.

A Tabela 48 mostra os níveis de colesterol de 12 participantes no início e no final do programa.

Programa		Diferença	Desvio	Desvio ao Quadrado
Início (x_1)	Final (x_2)	$d = x_1 - x_2$	$d - \bar{d}$	$(d - \bar{d})^2$
201	200	1	-19,16	367,36
231	236	-5	-25,16	633,36
221	216	+5	-15,16	230,03
260	233	+27	6,83	46,69
228	224	+4	-16,16	261,36
237	216	+21	0,83	0,69
326	296	+30	9,83	96,69
235	195	+40	19,83	393,36
240	207	+33	12,83	164,69
267	247	+20	-0,16	0,03
284	210	+74	53,83	2898,03
201	209	-8	-28,16	793,36
$\bar{x}_1 = 244,25$	$\bar{x}_2 = 224,08$	$\bar{d} = 20,17$	—	$\sum = 5885,7$

Tabela 48: Níveis de colesterol no início e no final do programa

As médias antes e depois do programa são 244,25 e 224,08, correspondendo a uma redução média no nível de colesterol de 20,17 com um desvio-padrão da redução de $s_d = \sqrt{5885,7/11} = 23,13$ e um desvio-padrão da redução média de $23,13/\sqrt{12} = 6,68$.

Apenas dois participantes tiveram o nível de colesterol aumentado após o programa, mas por pequenas quantidades (5 e 8 mg/dL).

A estatística de teste é:

$$T_p = \frac{20,17}{6,68} = 3,02 (p = 0,012)$$

Este valor-p é obtido da distribuição t de Student com 11 graus de liberdade. Concluímos que há evidência de que, em média, o programa altera o nível de colesterol.

9.6 Testes Não-Paramétricos

Os testes t apresentados nas seções anteriores foram derivados para situações em que o efeito do tratamento é descrito por uma variável contínua com distribuição gaussiana. Estes testes

fazem parte da classe de testes denominados *paramétricos*, pois baseiam-se nas estimativas dos parâmetros da distribuição assumida (neste caso, a distribuição gaussiana).

Os testes não-paramétricos são boas opções para situações em que ocorrem violações dos pressupostos básicos necessários para a aplicação de um teste paramétrico. Por exemplo, para testar a diferença de dois grupos quando a distribuição subjacente é assimétrica ou os dados foram coletados em uma escala ordinal. A distribuição da variável de interesse não é conhecida ou tem comportamento não gaussiano.

Dois destes testes não-paramétricos são apresentados neste material. O primeiro teste (teste de Mann-Whitney) é adequado para amostras independentes e o segundo (teste de Wilcoxon) para amostras pareadas.

Esses testes são construídos usando-se os postos das observações:

- O posto (rank) de uma observação é o número de ordem da observação, estando as observações ordenadas.
- Quando há empates, toma-se como posto de cada observação a média dos postos que seriam atribuídos às observações caso os empates não existissem.

9.6.1 Teste de Mann-Whitney

O teste de Mann-Whitney é usado para a comparação de dois grupos independentes.

Para sua construção:

1. seja n_1 o tamanho de amostra do menor dos dois grupos e n_2 o tamanho de amostra do maior dos dois grupos;
2. obtemos os postos de todas as observações como se os dois grupos fossem uma única amostra;
3. calculamos a estatística de teste
$$MW = n_1 n_2 + \frac{n_1(n_1+1)}{2} - T$$
em que n_1 e n_2 são os tamanhos das amostras dos dois grupos e T é a soma dos postos do grupo menor.

Para a tomada de decisão o valor da estatística MW pode ser comparado com o percentil de uma distribuição especial, ou podemos usar o resultado de que para estudos com pelo menos 10 observações em cada grupo T tem aproximadamente distribuição gaussiana com média
$$\mu_T = \frac{n_1(n_1+n_2+1)}{2}$$
e desvio-padrão
$$\sigma_T = \sqrt{\frac{n_2\mu_T}{6}}.$$

Neste caso, o valor da estatística $Z = \frac{T-\mu_T}{\sigma_T}$ deve ser comparado com o percentil da distribuição gaussiana padrão.

Exemplo: Comparação da tianeptina com placebo (continuação)

Comparando o grupo que recebeu tianeptina com o que recebeu placebo através do teste de Mann-Whitney temos a seguinte tabela com os dados ordenados e os postos correspondentes:

Placebo		Tianeptina	
Escore	Posto	Escore	Posto
2	1,5	2	1,5
		3	3
		4 e 4	4,5
6	6		
7	7,5	7	7,5
		8	9
		9	10
10	12	10 e 10	12
13	14,5	13	14,5
		14 e 14	16,5
15	18		
		17 e 17	19,5
21 e 21	22	21	22
26 e 26	24,5		
29 e 29	27	29	27
33 e 33	29,5		
37	31		

Tabela 49: Postos atribuídos aos dados agregados dos dois grupos (placebo e tianeptina)

Temos então que: $n_1 = 15$, $n_2 = 16$, $T = 1,5 + 6 + \cdots + 31 = 296,5$, $\mu_T = \frac{15(15+16+1)}{2} = 240$ e $\sigma_T = \sqrt{\frac{16\times 240}{6}} = 25,3$.

$$Z = \frac{T - \mu_T}{\sigma_T} = \frac{296,5 - 240}{25,3} = -2,2 (p = 0,027)$$

O valor-p indica que há diferença entre os dois grupos comparados. Este resultado é muito maior do que o valor-p obtido pelo teste t de Student ($p = 0,0104$).

Observe que o uso de um teste não-paramétrico é mais apropriado já que a escala usada para medir depressão é ordinal.

9.6.2 Teste de Wilcoxon

O teste de Wilcoxon é usado para comparar dois tratamentos quando os dados são obtidos através do esquema de pareamento.

Os seguintes passos devem ser seguidos na sua construção:

1. Calcular a diferença entre as observações para cada par.

2. Ignorar os sinais das diferenças e atribuir postos a elas.

3. Calcular a soma dos postos (S) de todas as diferenças negativas (ou positivas).

Para amostras pequenas (até 25 pares), o valor-p deve ser obtido através de uma tabela especial.

Para amostras grandes, a estatística do teste (S) tem aproximadamente distribuição gaussiana com média

$$\mu_S = \frac{n(n+1)}{4}$$

e desvio-padrão

$$\sigma_S = \sqrt{\frac{n(n+1)(2n+1)}{24}}$$

Assim, o valor de

$$Z = \frac{S - \mu_S}{\sigma_S}$$

deve ser comparado ao valor do percentil da distribuição gaussiana.

Exemplo: Evolução do tratamento com tianeptina

Rocha (1995) verificou se houve diminuição do escore de depressão entre os pacientes de um dos grupos durante o desenvolvimento do estudo.

A Tabela 50 mostra os escores dos pacientes do grupo tianeptina que foram admitidos em Belo Horizonte no primeiro dia (x_1) e no último dia (x_{42}).

No.	x_1	x_{42}	$d = x_{42} - x_1$	No.	x_1	x_{42}	$d = x_{42} - x_1$
1	24	6	-18	9	35	37	+2
2	46	33	-13	10	30	15	-15
3	26	21	-5	11	38	2	-36
4	44	26	-18	12	38	21	-17
5	27	10	-17	13	31	7	-24
6	34	29	-5	14	27	*	*
7	33	33	0	15	34	*	*
8	25	29	+4	16	32	26	-6

Tabela 50: Escores do grupo tianeptina no primeiro e último dias

Como o escore de MADRS é, na realidade, uma medida em escala ordinal, o uso de teste não-paramétrico é mais apropriado.

Dois pacientes (14 e 15) não completaram o estudo. Como não fornecem informações suficientes para a comparação pretendida, foram excluídos da análise.

As diferenças em valor absoluto, $|d|$, e os respectivos postos estão mostrados na Tabela 51.

$S = 3$, $\mu_S = 13(13+1)/4 = 45,5$, $\sigma_S = \sqrt{\frac{13(13+1)(26+1)}{24}} = 14,31$

$\|d\|$	Posto	No. paciente	$\|d\|$	Posto	No. paciente
0	-	7	14	7	10
2	1	9	16	8,5	5
4	2	8	16	8,5	12
5	3,5	3	17	10,5	1
5	3,5	6	17	10,5	4
6	5	16	23	12	13
13	6	2	35	13	11

Tabela 51: Posto atribuído ao valor absoluto da diferença $|d|$

$Z = (S - \mu_S)/\sigma_S = (3 - 45,5)/14,31 = -2,97(p = 0,003)$

Este valor-p é maior que o obtido através do teste t ($p = 0,0013$). Há portanto indicação de alteração dos níveis de depressão para pacientes que fizeram uso da tianeptina.