

Odival Faccenda
Lourdes Lago Stefanelo

Bioestatística Aplicada
ao Movimento Humano

Dourados-2014

Copyright © 2014
Odival Faccenda, Lourdes Lago Stefanelo

Revisão
Ângela Faccenda
José Roberto de Almeida

Capa
Margarida Langer

Diagramação
Edson Pereira de Souza

Ficha catalográfica elaborada pela Biblioteca Central da UEMS

Faccenda, Odival; Stefanelo, Lourdes Lago
Bioestatística aplicada ao movimento humano/Odival Faccenda,
Lourdes Lago Stefanelo. Dourados, MS: UEMS, 2014. 116p. ; 21cm.

Bibliografia
ISBN on-line: 978-85-99880-68-5

1.Bioestatística 2.Agilidade 3.Ritmo. I.Título.

CDD 20.ed. 574.015195

Direitos desta edição reservados aos autores.
É proibida a reprodução total ou parcial desta obra sem autorização dos autores.

UNIVERSIDADE ESTADUAL DE MATO GROSSO DO SUL

Reitor

Prof. Dr. Fábio Edir dos Santos Costa

Vice-Reitora

Prof^a. Dr^a. Eleuza Ferreira Lima

Pró-Reitor de Extensão, Cultura e Assuntos Comunitários

Prof. Dr. Edmilson de Souza

Divisão de Publicações

Prof^a. Me. Luíza Mello Vasconcelos

Conselho Editorial

Presidente

Prof^a. Dr^a. Márcia Maria de Medeiros

Editora

Prof^a. Me. Eliane Souza de Carvalho

Conselheiros

Prof. Dr. Edmilson de Souza

Prof^a. Me. Luíza Mello Vasconcelos

Prof. Dr. Luís Humberto da Cunha Andrade

Prof. Me. Esmael Almeida Machado

Prof. Dr. Hamilton Kikuti

Prof^a. Dr^a. Lourdes Missio

Prof^a. Dr^a. Silvane Aparecida de Freitas

Prof. Dr. William Fernando Antonialli Junior



Bloco A – Cidade Universitária

Caixa Postal 351 – CEP 79.804-970 – Dourados-MS

Fone: (67) 3902-2698

editorauems@uems.br

www.uems.br/site/editora



Odival Faccenda possui graduação em Matemática Licenciatura Plena pela Faculdade Porto Alergrense de Educação Ciências e Letras (1979), mestrado em Estatística pela Universidade Estadual de Campinas (1982) e doutorado em Agronomia (Energia na Agricultura) pela Universidade Estadual Paulista Júlio de Mesquita Filho (1997). Pós-doutorado em educação pela Unicamp (2011). Atualmente é professor titular da Universidade Estadual de Mato Grosso do Sul. Tem experiência

na área de Matemática e Estatística Aplicada, com ênfase em Energia na Agricultura, atuando principalmente nos seguintes temas: alelopatia, avaliação ambiental, avaliação educacional, avaliação e monitoramento das necessidades sociais e de saúde de idosos, saúde mental, controle da qualidade de processos dinâmicos em tempo real, biodiversidade, controle biológico, manejo e conservação de insetos.



Lourdes Lago Stefanelo possui graduação em Educação Física pela Universidade de Cruz Alta (1977), Especialização em Administração de Recursos Humanos, Especialização em Ginástica, Especialização em Educação Psicomotora (pela Universidade de Cruz Alta), Especialização em Planejamento Energético Ambiental dos Municípios do Rio Grande do Sul pela UFRGS, mestrado em Educação Física pela Universidade

de São Paulo (1989), doutorado em Administração Educacional - University of Wisconsin - Madison (2001). Doutorado em Ciências da Educação - Universidad Técnica de Comercialización e Desarrollo - UTCD, Assunção/PY. Atualmente é prof^a. assistente da Universidade Estadual de Mato Grosso do Sul, prof^a. de Introdução à Metodologia Científica nos cursos de Física, Matemática, Enfermagem e Ciência da Computação, coord^a. do curso de Educação Física - PARFOR. Tem experiência na área de Educação, com ênfase como Coordenadora do Centro de Extensão da Unicruz/Cruz Alta, RS, 1990/92, Assessora de Assuntos Comunitários 1993/94; como pesquisadora atuando principalmente nos seguintes temas: educação ambiental, educação, avaliação física e postural, educação física adaptada, aptidão física e atividade física. Organizadora dos Livros: Sessões Historiadas: a criatividade na escola e Novas Histórias para Encenação na Escola. Editora da Revista Reflexos da Ciência.

APRESENTAÇÃO

A estatística revolucionou a ciência, durante o século XX, sofisticando o processo de pesquisa na direção de melhores parâmetros de investigação e permitindo, assim, tomadas de decisões mais sábias, nas diversas áreas do conhecimento humano. No entanto, há entre nós, brasileiros, certa resistência contra a utilização e a compreensão dos números nas questões biológicas, sobretudo, por parte dos profissionais da saúde.

Não se sabe muito bem a razão desta evitação, afinal, os números sempre educaram a humanidade. Começamos a contar, medir, enumerar há milhares de anos, e esta situação continua sendo uma necessidade. Os números detêm a essência da lógica, têm estabilidade, são independentes de opiniões, classificações, conceitos e preconceitos pessoais e, assim que são reconhecidos, reinam absolutos no conhecimento humano.

Todos os que gostam de conhecer, de saber, de investigar devem e podem ser amigos dos números. Sabe-se que quanto mais avançada uma área, mais quantificada é. Por isso é importante para qualquer profissional, principalmente os da saúde ter treinamento suficiente para usar adequadamente o arsenal de números que é oferecido pela Estatística.

As pesquisas na área da Educação Física e do Esporte, assim como em outras áreas da saúde, têm a intenção de melhorar nosso entendimento de como o corpo humano responde e se adapta ao movimento. Para que elas sejam confiáveis precisam ser bem planejadas e analisadas. A metodologia estatística é um instrumento para se atingir o objetivo, isto é, para examinar a questão da pesquisa ou de uma teoria.

Os avanços tecnológicos em computação criaram “pacotes” para análises estatísticas que permitem analisar grandes quantidades de

dados e realizar cálculos complexos, em um tempo muito pequeno, mas não oferecem (ainda) a reflexão de “como” e dos “porquês” de seus resultados. Assim, é importante reconhecer que a capacidade de interpretar os resultados dos pacotes estatísticos é uma exigência vital para uma investigação séria, e isso se consegue com o conhecimento básico, fundamental da Estatística.

Este é um livro que explica e esclarece, sem oprimir o leitor com excessos e detalhes, utiliza exemplos relacionados ao impressionável e infinito movimento humano.

Com toda certeza, é um livro necessário para todos aqueles que se iniciam na vida científica.

Prof. Dr. Valdir Barbanti
Prof. Titular da Escola de Educação Física e Esporte da USP

PREFÁCIO

Neste livro, pretendemos apresentar os conteúdos introdutórios de estatística aplicados aos ensaios do movimento humano. O objetivo é apresentar as ferramentas estatísticas como recursos auxiliares de análise e interpretação dos resultados de pesquisas executadas por alunos no decorrer do curso. Pretende-se, com isso, apresentar subsídios ao leitor, com técnicas de análise que possam auxiliar na interpretação, organização e apresentação dos dados provenientes de seus estudos. Com esse propósito, utilizaram-se dois exemplos de pesquisa que podem servir de apoio no direcionamento de atividades correlatas.

O primeiro refere-se ao estudo “Brincar, Coreografar e Perceber”, desenvolvido em pré-adolescentes, com 11 e 12 anos de idade, que será usado em atividades lúdicas básicas e específicas e de participação, na criação e elaboração de elementos coreográficos, utilizando-se a música, mesmo nas atividades de ginástica e de expressão corporal, para provocar *insights*, especificamente, na percepção do ritmo e do entendimento das partes do próprio corpo e o de outras pessoas, ampliando as percepções corporal, espacial, direcional e temporal.

É natural que dessas atividades decorram indagações, tais como: essas e outras atividades alternativas a serem propostas e desenvolvidas com música despertam prazer e o interesse de se realizá-las? Tais atividades melhoram a noção de lateralidade e ritmo dos pré-adolescentes?

Nesse contexto, e com o propósito de responder a tais questionamentos, foi planejada e executada uma pesquisa com pré-adolescentes, em escolas da Rede Municipal de Ensino de Dourados/MS e de Jardim/MS, que teve como objetivo verificar se existiam assimetrias bilaterais de membros inferiores e de ritmo, cujos resultados estão apresentados no Apêndice A. A interpretação desses resultados encontra-se em Stefanelo et al, (2011), que recomenda aos seus autores

que voltem ao grupo e ofereçam uma programação de atividades básicas e criativas que desenvolvam as habilidades de exploração motora, de ritmo, para verificar se elas beneficiam o autoconceito dos pré-adolescentes.

O outro exemplo de pesquisa refere-se à Escala de Sono de Epworth, versão em português do Brasil (ESS/BR). A Escala de Sonolência de Epworth é um instrumento empregado em estudos clínicos e pesquisas, em vários países, incluindo o Brasil. Ela quantifica a chance de um indivíduo cochilar em oito situações diferentes. Possui alta sensibilidade e confiabilidade, além de elevado grau de consistência interna, podendo discriminar pacientes com níveis diferentes de sonolência e indivíduos normais, além de detectar variações de Sonolência Excessiva Diurna (SED) em pacientes submetidos a tratamento, especialmente nos portadores de apneia obstrutiva do sono. Neste estudo buscou-se identificar estratos da população de Jardim/MS que representasse risco em relação à sonolência, com o intuito de estudar práticas que auxiliem na melhora do problema. Os resultados estão apresentados no Apêndice B.

O livro foi escrito em capítulos, cada um deles apresentando um aspecto da teoria estatística elementar. Os exemplos e aplicações dos conceitos estão relacionados com o tema “ensaios do movimento humano”, com os quais pretendemos mostrar ao leitor a importância das técnicas estatísticas na interpretação de temas relacionados com as práticas de Educação Física. Os exemplos de aplicação de conceitos devem ser lidos com cuidado, mantendo-se o espírito investigativo. Desta forma, sugerimos que o estudante defina um problema a ser investigado para trabalhar os temas individualmente, sob a orientação de um professor.

Acreditamos que o texto possa ser utilizado em disciplinas de graduação, e oferecer recursos valiosos nos trabalhos de pesquisa, de iniciação científica e de conclusão de curso. Na resolução dos exercícios é apresentado um roteiro orientador, sobre o uso de aplicativos computacionais como ferramenta de auxílio na resolução dos exercícios.

No capítulo 1, apresentamos as técnicas da estatística descritiva, juntamente com a análise exploratória dos dados da pesquisa sobre lateralidade e ritmo. Nessa discussão, abordamos a natureza da estatística, o tratamento da informação (classificação de variáveis e níveis de mensuração), as distribuições de frequência, os gráficos e as medidas-resumo (posição e dispersão).

Em seguida, estudamos noções sobre probabilidade e variáveis Aleatórias e apresentamos os pressupostos minimamente essenciais para que o leitor não familiarizado com a estatística possa entender os conceitos de inferência e teste de hipóteses, usados na interpretação dos resultados da pesquisa.

No capítulo seguinte, são apresentados ao leitor, de forma resumida, os conceitos de estimador, intervalo de confiança para um parâmetro, regra de decisão, erros Tipo I e II e principais estimadores usados no julgamento de hipóteses.

No capítulo final, estudamos, por meio de exemplos, a melhor maneira de se formularem hipóteses estatísticas a partir de dados coletados e quais devem ser o teste apropriado e a melhor forma de se apresentarem os resultados.

Enfim, podemos resumir a intenção deste livro como uma tentativa de se obter informação a partir de dados. Tentamos nos convencer de que nossa hipótese é correta ou não. Finalmente, tentamos convencer os terceiros, por meio de publicações, de que nossos resultados estão corretos. Entretanto, a aquisição das informações pode envolver conceitos vagos, imprecisões e contornos mal definidos que podem trazer dúvidas e incertezas à análise dos resultados. Assim, é fundamental, na hora de concluir, pensar sempre em todas as possibilidades de erro nos procedimentos metodológicos escolhidos e revelar os problemas identificados de forma transparente.

Somos gratos a todas as pessoas que colaboraram de alguma forma na execução dos testes e que contribuíram para a elaboração deste livro. Entre elas, citamos os professores Ednéia Albino Nunes Cerchiar; Marcos Antonio Araujo dos Santos, Marcia Cristina Bortoleto Rotta,

Maria Margarida Lago Stefanello e os alunos Adriney José da Silva, Alessandra Cavalheiro Vieira, Alexandre Florenciano da Silva, Allyson Eckert Cavalheiro, André Luís de Mello Martins, Edilaine Menezes da Cunha, Edilene de Moraes, Edymar Rocha Viedes, Flaviano de Oliveira Mendonza, George Luís dos Santos Peixoto, Gladis Peres Alves, Jair Cuevas, José Renato Ortiz de Arruda, Katia Regina Farias de Souza, Laura Cristiane F.Barros, Leonel Vieira, Lourival Franco, Márcia Correa Bueno, Márcio Roberto de B. Peixoto, Marcos Luciano Munhoz, Maryeli Cavalheiro de Campos, Milena dos Santos Peixoto, Rosa Mara Aguirres Rolão, Rosana Arguello Marques, Simone Regina Nienkotter, Vandira Santos da Cruz Lombardi, Wolney Rodrigues Vasques. Nossos agradecimentos extensivos às Direções e aos alunos das Escolas Municipais de Dourados Joaquim Murtinho, Aurora Pedroso, Clarice Bastos, Franklin Azambuja, Etalívio Penzo e de Jardim, Oswaldo Fernandes Monteiro, pela disposição em colaborar, relativamente à amostra de pessoas que fizeram parte deste estudo. Ao Ministério da Educação pelo apoio, obrigado.

Dourados, março de 2014.

Odival Faccenda
Lourdes Lago Stefanello

SUMÁRIO

CAPÍTULO I	21
1. ANÁLISE EXPLORATÓRIA DE DADOS.....	21
1.1. INTRODUÇÃO.....	21
1.2 PLANEJAMENTO DA COLETA DE DADOS	22
1.3 DEFINIÇÕES	23
1.4. TIPOS DE VARIÁVEIS	25
1.5 RELAÇÃO DE DEPENDÊNCIA ENTRE AS VARIÁVEIS.....	27
1.6 ELABORAÇÃO DA PLANILHA COM OS REGISTROS.....	31
1.6.1 Como criar um banco de dados.....	32
1.6.2 Crítica dos dados para detectar observações atípicas	34
1.7 APRESENTAÇÃO DOS DADOS DAS VARIÁVEIS EM TABELAS	34
1.7.1 Classificação das séries estatísticas.....	37
Exercícios suplementares.....	38
1.7.2 Tabelas de dupla entrada	40
Exercícios:.....	41
1.8 GRÁFICOS.....	41
1.8.1 GRÁFICO SEQUENCIAL, GRÁFICO DE TENDÊNCIA OU DIA-GRAMA POR LINHA POLIGONAL	41
1.8.2 GRÁFICO DE COLUNAS.....	43
1.8.3 GRÁFICO EM SETORES.....	44
1.9 DISTRIBUIÇÃO DE FREQUÊNCIA.....	45
1.9.1 REPRESENTAÇÃO DA AMOSTRA PARA VARIÁVEL QUANTITATIVA CONTÍNUA	45
1.9.2 REPRESENTAÇÃO DA AMOSTRA PARA VARIÁVEL QUANTITATIVA INTEIRA.	48
1.10 MEDIDAS UNIVARIADAS	49
1.11 MEDIDAS DE TENDÊNCIA CENTRAL PARA UMA AMOSTRA....	49

1.11.1 MÉDIAS	49
1.11.2 MEDIANA.....	52
1.11.3 MODA (Mo)	53
1.12 MEDIDAS DE DISPERSÃO OU VARIABILIDADE	55
1.12.2 VARIÂNCIA (σ^2)	56
1.12.3 DESVIO PADRÃO (σ_x OU S_x).....	56
1.12.4 MEDIDAS DE ACHATAMENTO OU CURTOSE	57
CAPÍTULO II.....	61
2. NOÇÕES SOBRE PROBABILIDADE E VARIÁVEIS ALEATÓRIAS.....	61
2.1 DEFINIÇÕES BÁSICAS	62
2.1.1 CARACTERIZAÇÃO DE UM EXPERIMENTO ALEATÓRIO NÃO DETERMINÍSTICO	62
2.1.2 ESPAÇO AMOSTRAL.....	63
2.1.3 EVENTOS	64
2.1.4 PROBABILIDADE	64
2.2 PROBABILIDADE CONDICIONAL.....	65
2.3 EVENTOS INDEPENDENTES.....	67
2.4 TEOREMAS	67
2.5 VARIÁVEL ALEATÓRIA.....	68
2.6 DISTRIBUIÇÃO DISCRETA	69
2.7 DISTRIBUIÇÃO BINOMIAL	70
2.8 DISTRIBUIÇÕES DE VARIÁVEIS ALEATÓRIAS CONTÍNUAS.....	72
2.9 DISTRIBUIÇÃO NORMAL REDUZIDA.....	74
2.10 COMO CALCULAR PROBABILIDADES NA DISTRIBUIÇÃO NORMAL.....	75
CAPÍTULO III.....	77
3. INFERÊNCIA E TESTE DE HIPÓTESE	77
3.1 DEFINIÇÕES E CONCEITOS BÁSICOS.....	77

3.2 NOÇÃO DE ESTIMADOR.....	79
3.2.1 DISTRIBUIÇÕES PARA AMOSTRAS.....	79
3.2.2 INTERVALO DE CONFIANÇA PARA A MÉDIA.....	83
3.3 REGRA DE DECISÃO E PROBABILIDADE DE ERRO TIPO I E TIPO II.....	85
3.3.1 PASSOS PARA CONSTRUÇÃO DE UM TESTE DE HIPÓTESES.....	88
3.4 ESTIMADORES OU ESTATÍSTICAS USADOS PARA JULGAR HIPÓTESES.....	90
4 FORMULAÇÃO DE HIPÓTESE ESTATÍSTICA, TESTE APROPRIADO, CAPÍTULO IV.....	94
FORMULAÇÃO DE HIPÓTESE ESTATÍSTICA, TESTE APROPRIADO, APRESENTAÇÃO DO RESULTADO.....	94
4.1 COMPARAR DUAS MÉDIAS COM PRESSUPOSTO DE NORMALIDADE.....	94
4.2 COMPARAR DUAS MÉDIAS SEM PRESSUPOSTO DE NORMALIDADE.....	96
4.3 ASSOCIAÇÃO ENTRE DUAS VARIÁVEIS CATEGÓRICAS.....	98
4.4 COMPARAR MAIS DE DUAS MÉDIAS COM PRESSUPOSTO DE NORMALIDADE.....	99
4.5 EXISTÊNCIA DE CORRELAÇÃO ENTRE DUAS VARIÁVEIS.....	101
4.6 EXISTÊNCIA DE UMA RELAÇÃO DE DEPENDÊNCIA ENTRE VARIÁVEIS.....	103
CONCLUSÃO.....	106
REFERÊNCIAS.....	107
Anexo A: TABELAS.....	110
Tabela de valores críticos – Normal.....	111
Tabela de valores críticos – t de Student.....	112
Tabela de valores críticos – Qui-quadrado.....	113

Tabela de valores críticos – F de Snedecor.....	114
Tabela de Números Aleatórios.....	115
Apêndices.....	116
Anexo B. Escala de sono de Epworth Versão em Português no Brasil (ESS/BR).....	114
A. Planilha Ritmo e Lateralidade.....	118
B. Planilha Escala de Sono.....	123

LISTA DE FIGURAS

Figura 1.4.1 Diagrama do relacionamento entre tipos de mensuração e suas características.	26
Figura 2.5.1 - Variável aleatória e função de probabilidade.	69
Figura 2.8.1 - Aproximação da distribuição normal.	73
Figura 2.8.2 - Distribuição normal.....	73
Figura 2.9.1 - Distribuição normal padrão.	74

LISTA DE QUADROS

Quadro 1.5.1 - Tipos de variáveis segundo a função no plano de análise.....	27
Quadro 1.5.2 - Exemplos de Relação entre plano de análise e tipo de variável.	28
Quadro 1.5.3 - Tipo de mensuração versus possíveis testes apropriados.....	29
Quadro 1.5.4 - Palavras que descrevem diferentes graduações, entre 0% e 100%, segundo escala de intensidade de 5 pontos.....	31
Quadro 3.3.2 - Resumo dos erros (estado) e as decisões.....	87

LISTA DE GRÁFICOS

Gráfico 1.8.1 - Número médio de desvios no teste de ritmo.....	42
Gráfico 1.8.2 - Frequência de tempo no teste de ritmo.....	43
Gráfico 1.8.3 - Percentual de alunos por classificação de acertos.....	44
Gráfico 1.9.1 - Distribuição de alunos por intervalo de peso.....	47
Gráfico 1.9.2 - Distribuição de alunos por intervalo de peso.....	47
Gráfico 1.9.3 - Número de alunos por quantidade de desvios no tempo 1.....	48
Gráfico 3.3.1 - Região crítica da distribuição normal.....	89
Gráfico 4.5.1 - Distribuição do tempo para execução do teste em função do peso do estudante.....	102
Gráfico 4.6.1 - Relação entre peso do estudante e tempo gasto para executar o teste das três faixas.....	105

LISTA DE TABELAS

Tabela 1.6.1 - Planilha resumo das medidas sobre o teste de lateralidade.....	33
Tabela 1.6.2 - Planilha resumo do teste ESS_BR (Escala de Sono).....	33
Tabela 1.7.1 - Frequência segundo a idade dos alunos da amostra sobre lateralidade, Dourados, 2010.....	37
Tabela 1.7.2 - Frequência por número de desvios no teste de deslocamento lateral T1 velocidade 12, dos alunos de Dourados, Apêndice A.....	39
Tabela 1.7.3 - Número de alunos por intervalo de peso dos alunos de Dourados que compõe a amostra do Apêndice A.....	39
Tabela 1.7.4 - Frequência e porcentagem nos testes de deslocamento lateral por categoria de sexo dos alunos, Dourados, 2010.....	40
Tabela 1.8.1 - Número médio de desvios da turma X, no teste de ritmo, ao longo de 4 anos, nas três velocidades.....	42
Tabela 1.8.2 - Frequência por intervalo de tempo para execução do teste de deslocamento, Dourados, 2010.....	43
Tabela 1.8.3 - Resultado do teste de ritmo.....	44

LISTA DE TABELAS (continuação)

Tabela 1.9.1 - Número de alunos por intervalo de peso submetidos ao teste da lateralidade.	46
Tabela 1.9.2 - Frequência de desvios no teste de ritmo no Tempo 1.....	48
Tabela 1.11.1 - Peso médio dos estudantes por ano de idade.	51
Tabela 1.11.2 - Notas e pesos de quatro provas do aluno Ricardo.....	51
Tabela 1.11.3 - Escala de tempo para execução do teste da agilidade (lateralidade).....	53
Tabela 1.12.1 - Notas de quatro alunos em quatro testes diferentes.....	55
Tabela 2.3.1 Resultados possíveis no lançamento simultâneo de uma moeda e um dado.	67
Tabela 2.6.1 - Distribuição probabilística do número de desvios do teste no T1.....	70
Tabela 2.7.1 - Distribuição do número de desvios em 5 alunos escolhidos aleatoriamente.	71
Tabela 3.2.1 - Distribuição da amostra de algumas estatísticas obtidas da amostra de tamanho $n = 2$, retiradas da população (1, 2, 3).....	80
Tabela 3.2.2 - Distribuição amostral de \bar{X}	80
Tabela 3.2.3 - Resumo de estimadores.	84
Tabela 3.3.1 - Valores de α e β para diferentes regiões críticas, com n fixo igual a 10.....	87
Tabela 4.1.1 - Distribuição das médias e demais medidas do tempo de execução do teste em relação ao gênero.	95
Tabela 4.2.1 - Distribuição das médias do número de erros e demais medidas desta variável em relação ao gênero, peso e altura.....	97
Tabela 4.3.1 - Frequência no teste T1 de ritmo, por sexo, peso e altura dos participantes e significância no teste.....	98
Tabela 4.4.1 - Distribuição das médias do tempo necessário para execução do Teste das três faixas de acordo com a idade do aluno.	100
Tabela 4. 6.1 - Relatório de saída do aplicativo.....	104

Capítulo 1

Análise Exploratória de Dados

Para estudar a motricidade humana é necessário que se tenha uma noção de ser humano, o que favorece a eclosão do conhecimento que pretendemos analisar. Também é preciso considerar a “metaproblemática da complexidade do homem” (MOREIRA, 1993), especificamente, na constatação de fatos que é a busca de respostas as nossas inquietações quanto às habilidades das pessoas, pois, conforme Fraisse e Piaget (1972), “não se encontra o que não se procura”. Neste texto procuramos combinar técnicas estatísticas com indagações sobre aspectos relacionados ao ser humano e, em particular, ao desenvolvimento motor de adolescentes, adquirindo habilidades na formulação de hipóteses, na definição das variáveis candidatas a elucidarem o problema e nas técnicas adequadas para uma análise descritiva e quantitativa na averiguação das hipóteses.

1.1 INTRODUÇÃO

A finalidade da análise estatística exploratória ou estatística descritiva é apresentar os principais conceitos sobre o levantamento de dados, destacar as técnicas de apresentação, por meio de tabelas e gráficos, oferecer as medidas próprias para análises e as técnicas usadas para a interpretação dos dados numéricos. Fonseca e Martins (1994) resumem o procedimento, no seguinte diagrama:



Fonte: FONSECA; MARTINS (1994).

É preciso considerar que os números não falam por si mesmos. Precisam ser organizados, discutidos e interpretados. Então, *a estatística pode ser vista como um conjunto de técnicas que ajudam na tomada de decisão, quando prevalecem condições de incerteza.*

1.2 PLANEJAMENTO DA COLETA DE DADOS

Em um trabalho de investigação, existe uma sequência a ser respeitada. Após identificar o assunto a ser estudado (percepção de uma lacuna no conjunto de conhecimentos) e de se estabelecerem o planejamento da pesquisa das variáveis relevantes a serem incluídas (quais são as perguntas que o investigador propõe? quais são as hipóteses a comprovar?), a forma pela qual os dados das variáveis serão coletados (descrever o método ou procedimento utilizado na obtenção do dado), o cronograma das atividades, os custos envolvidos (é pertinente gastar recursos para tentar conseguir respostas a tais perguntas?), o exame das informações disponíveis (o que já se sabe ou foi publicado sobre o assunto), e o delineamento da amostra, o passo seguinte é a coleta dos dados, que consiste na busca ou compilação dos dados das variáveis componentes do fenômeno a ser estudado. Exemplos de coleta de dados que iremos estudar neste livro são Ritmo e Lateralidade, Apêndice A, como exemplo para as atividades práticas no decorrer dos capítulos, e a Escala de Sonolência, Apêndice B, que deverá servir de aplicação complementar aos conteúdos apresentados nos diferentes capítulos do livro.

Atividade 1. Providenciar o formulário preenchido de alguma atividade prática relacionada com a sua especificidade.

Os subsídios para esta etapa da pesquisa geralmente podem ser encontrados em tópicos de metodologia científica (GOLDENBERG et al. 1991). De maneira geral, o planejamento da pesquisa tem início com algumas indagações sobre: o que investigar? O que procurar? Qual o tema? Qual a abrangência do tema que se irá investigar? Essas indagações podem se reportar a comportamentos, qualidade de vida, sono, habilidades, coordenação motora, rendimentos, valor

acrescentado, performance, saúde, enfim, a qualquer assunto que possa ter algum interesse prático. Suponha-se que o tema escolhido seja habilidades e, dentro desse tema, pretenda-se investigar a habilidade de deslocar-se lateralmente ou fazer o teste de ritmo. Podemos observar que é necessário a percepção do andamento do ritmo. Para executar o deslocamento depende de outras variáveis tais como gênero, idade, desenvolvimento, agilidade, ritmo, entre outros. Outra exemplificação, referente ao tema performance, é o rendimento de atletas que disputam a modalidade dos 100 metros rasos. A variável resposta que se tem interesse é o tempo em segundos que o atleta leva para fazer o percurso, mas sabe-se que existem variáveis que interferem nesse tempo, tais como idade, estatura, genética, tempo de treinamento e orientação nutricional, entre outras. Outro aspecto que dificulta o pesquisador é que algumas dessas variáveis não são passíveis de mensuração direta e dependem de testes de laboratório, por isso não são facilmente identificadas em escolas de ensino básico ou médio, apenas em laboratórios de clubes, universidades e hospitais, por exemplo.

1.3 DEFINIÇÕES

Antes de abordar as medidas e a melhor forma de analisá-las, apresentamos alguns conceitos que geralmente são utilizados e, às vezes, geram alguma confusão. Define-se **população** como o conjunto de elementos que têm, em comum, determinada característica, ou seja, é o conjunto de elementos mantidos sob as mesmas condições, sejam elas naturais ou artificiais. No exemplo, Apêndice A, a população é o conjunto de todos os alunos pré-adolescentes, na faixa etária dos 9 aos 11 anos, das escolas municipais de Dourados.

Variáveis são as características ou os caracteres mensuráveis que se quer estudar e que exibem certo grau de variabilidade, quando submetidas a mensurações sucessivas. No exemplo, são todas as sentenças que encabeçam as colunas da planilha. É importante fazer a definição conceitual de cada variável. Imagine que se pretenda saber a proporção de toxicodependente existente em uma determinada escola. Então, precisamos definir o que é ser toxicodependente. O fumante é toxicodependente? Ou, apenas o consumidor de drogas ilícitas é toxicodependente?

Amostra. Todo subconjunto (porção) considerado aleatoriamente de uma população é denominado uma amostra dessa população. É usada, em geral, para se testarem hipóteses ou fazer inferência sobre algum parâmetro desconhecido que se pretenda conhecer. O termo aleatório significa que foi feito um sorteio para a escolha dos elementos do subconjunto.

O termo **indivíduo** pode designar, segundo o caso, o empregado de uma empresa, um aluno, um animal, uma escola, um professor etc. É sempre a entidade de base, sobre a qual o observador realiza um certo número de medições. No exemplo, Apêndice A, é o estudante pré-adolescente.

Recenseamento é quando a coleta das informações incide sobre todos os elementos da população. Quando são coletadas informações de apenas parte da população, diz-se que foi feita uma **amostragem**.

Os pesquisadores trabalham com amostras por vários motivos. Um deles é o fato de que as populações infinitas só podem ser estudadas por intermédio de amostras. Por exemplo, por maior que seja o número de vezes que um pesquisador pese um corpo sólido, o resultado será sempre uma amostra porque, teoricamente, todo o corpo pode ser pesado um número infinito de vezes.

Outro fator, as populações finitas muito grandes só podem ser estudadas através de amostras. Por exemplo, o número de cobaias existentes no mundo, em determinado período de tempo, é matematicamente finito, mas as pesquisas que usam cobaias só podem ser feitas com amostras, porque nenhum pesquisador dispõe de todas as cobaias do mundo para seu trabalho.

Em um **estudo descritivo** é suficiente fazer a análise de forma isolada para cada variável: é a **análise uni-variada**. No **estudo analítico**, as variáveis estão implícitas nas hipóteses que serão verificadas quanto a sua veracidade ou não: é a análise **bi ou multivariada**.

1.4 TIPOS DE VARIÁVEIS

Como exemplo, vide o Apêndice B sobre a Escala de Sonolência. Vemos ali que o pesquisador colheu informações sobre idade, sexo e mais oito variáveis: V1, V2, ..., V8. Poder-se-iam ter incluído outras: como profissão, cor, estado civil etc. No exemplo, no Apêndice A, sobre ritmo e lateralidade, o pesquisador escolheu idade, peso, altura, ano escolar, T1, T2, T3, Tempo de execução, lado que iniciou e reiniciou o teste. Além dessas, encontram-se outras informações na página anexa ao formulário preenchido.

Algumas variáveis, como sexo, deficiência e escolaridade apresentam, como possíveis realizações, uma qualidade (ou atributo) do indivíduo pesquisado, ao passo que outras, como peso, número de desvios e idade apresentam números resultantes de uma contagem ou mensuração, como possíveis realizações. As variáveis do primeiro tipo são chamadas de **qualitativas** e as do segundo tipo são chamadas de **quantitativas**.

Dentro das variáveis qualitativas, pode-se ainda fazer distinção entre dois tipos: **variável qualitativa nominal**, para a qual não existe qualquer ordenação nas possíveis realizações, e **variável qualitativa ordinal**, para a qual existe uma certa ordem nos possíveis resultados. Nos exemplos, sexo é um caso de variável nominal, ao passo que a escolaridade é um exemplo de variável ordinal, pois 4º ano, 5º ano e 6º ano correspondem a uma ordenação baseada no número de anos de escolaridade.

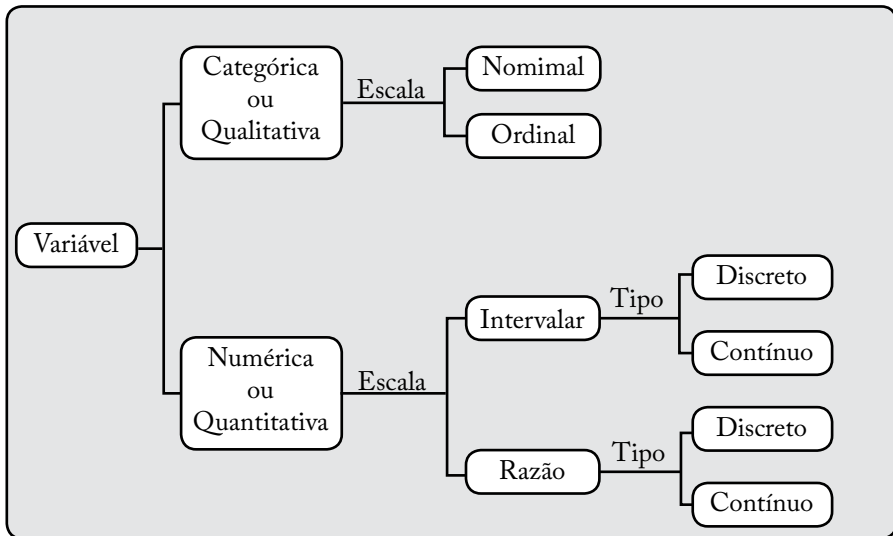
De modo análogo, as variáveis quantitativas apresentam duas classificações, em **escala de intervalo** ou em **escala de razão**, e elas são classificadas como discretas e contínuas, sendo, **variável quantitativa discreta** quando suas possíveis realizações assumem somente valores inteiros. É usada, em geral, para se fazer contagem. Exemplo: número de vezes que se desvia do ritmo. **Variável quantitativa contínua**: quando suas possíveis realizações podem assumir uma infinidade de valores em um intervalo qualquer, por menor que seja. É usada, em geral, em medições. Exemplo: peso da pessoa, altura ou idade dentre outras.

As variáveis conhecidas como escala intervalar são aquelas em que o zero existe por convenção. Exemplo: temperatura (Celsius,

Fahrenheit); altura (metro, centímetro, pés); receitas de vendas (reais ou dólares); tempo (calendário gregoriano e hebraico). Já na **escala da razão** ou **proporcionalidade** existe um ponto zero verdadeiro, como origem, nesse caso, todas as operações aritméticas podem ser feitas. Exemplos: pressão, volume, massa, distância e outras.

Resumindo, a classificação das variáveis apresenta-se de acordo com o esquema da Figura 1.4.1.

Figura 1.4.1 Diagrama do relacionamento entre tipos de mensuração e suas características.



Fonte: os autores.

1.5 RELAÇÃO DE DEPENDÊNCIA ENTRE AS VARIÁVEIS

Ao escolher as medidas para as variáveis, é conveniente que o pesquisador considere ainda sua classificação, em relação ao plano de análise que pretende executar. Nessa perspectiva, podemos distinguir três tipos de variáveis, conforme a Quadro 1.5.1

Quadro 1.5.1 - Tipos de variáveis segundo a função no plano de análise

Tipo de variável	Característica
Dependente ou Resultado	Mede o fenômeno que se estuda e que se quer explicar. No exemplo, Desvio do Ritmo, que poderia ser medido por categoria (nominal – Sim/Não; Ordinal – Desvio leve, moderado, grave) ou Tempo de Execução do Teste das Três Faixas.
Independente ou Exposição	É (ou são) a(s) variável(eis) que se considera(m) candidata(s) a explicar a dependente. Ex.: idade, que poderia ser medida categoricamente (nominal – não defasado, defasado; ordinal – por ano de escolarização) ou por medida contínua (em anos meses e dias).
Interferente, ou de corte, ou, covariável	É (ou são) a(s) variável(eis) que se considera(m) capaz(es) de interferir na relação entre dependente e independente, podendo sugerir relações falsas que dizem respeito a sua interferência e não à relação estudada. Ex.: sexo, é uma variável em que a relação entre tempo de execução do teste e idade do aluno pode variar de maneira distinta.

Fonte: os autores.

As possibilidades de análise de tais variáveis estão ligadas às características de sua natureza, como qualitativa ou quantitativa. Daí, para a concepção (formulação) de variáveis, pode-se agregar aos cuidados na escolha de medidas também a consideração das possibilidades analíticas do tipo de medida a ser escolhida. O pesquisador deverá consultar os pré-requisitos da técnica analítica que pretende utilizar. O Quadro 1.5.2 ilustra algumas relações entre plano de análise e tipo de variável, com alguns exemplos.

Quadro 1.5.2 - Exemplos de Relação entre plano de análise e tipo de variável.

Plano de análise	Variável dependente	Variável independente	Variável de controle
Correlação (r Pearson)	Quantitativa contínua	Quantitativa contínua	Qualitativa (nominal ou ordinal) aplicada em correlações parciais.
Regressão linear	Quantitativa contínua	Preferencialmente quantitativa (embora não necessariamente)	Preferencialmente quantitativa, distinguindo diferentes posições paralelas de retas de regressão entre dependente e independente.
Associação em tabela de contingência (χ^2)	Qualitativa (nominal ou ordinal)	Qualitativa (nominal ou ordinal)	Qualitativa (nominal ou ordinal), criando estratificação para o contingenciamento.

Fonte: Siegel (1981); Levin (1985); Miller (1991).

No Quadro 1.5.3 são apresentados os tipos de mensuração com tipo de prova e possíveis testes estatísticos apropriados, úteis no momento de se definirem os testes a serem aplicados, tendo-se em mente as hipóteses do estudo e a escala de medição.

Quadro 1.5.3 - Tipo de mensuração versus possíveis testes apropriados.

Escala	Relações Definidoras	Tipos de Provas Estatísticas Adequadas	Medidas e Testes Estatísticos Adequados
Nominal	Equivalência	Não paramétricas	Número de Casos, Percentagens, Moda, Correlações de Contingência Qui-quadrado, Teste exato de Fisher, Frequência, Teste binomial, Mcnemar, Cochran Q_c , Coeficiente de Contingência, Moda.
Ordinal	Equivalência Maior do que	Não paramétricas	Todos os anteriores e mais: r_s de Spearman, τ de Kendal, W de Kendall, Mediana, Percentis, Decis, Quartis, Desvio Quartílico, Correlação de Postos, Teste de sinais, Teste de Kolmogorov-Smirnov, Teste U, Mann-Whitney, Teste de Kruskal Wallis.

Fonte: Siegel (1981); Levin (1985); Miller (1991).

Continuação do Quadro 1.5.3

Escala	Relações Definidoras	Tipo de Provas Estatísticas Adequadas	Medidas e Testes Estatísticos Adequados
Intervalar	Equivalência Maior do que Razão conhecida de dois intervalos quaisquer	Paramétricas e Não paramétricas	Todos os anteriores e mais: Média. Desvio-padrão, Média aritmética, Desvio padrão, variância, Desvio médio, Intervalo, Amplitude total, Amplitude média, Correlação de produto-momento (product-moment correlation), Teste-t, Teste-F, Teste-Z, ANOVA.
Razão	Equivalência Maior do que Razão conhecida de dois intervalos quaisquer Razão conhecida de dois valores quaisquer de escala	Paramétricas e Não paramétricas	Todos os anteriores e mais: Coeficiente de variação, Média geométrica, Média harmônica.

Fonte: Siegel (1981); Levin (1985); Miller (1991).

No Quadro 1.5.4 são apresentadas palavras utilizadas em questões que envolvem escalas graduadas, do tipo Likert de cinco pontos, úteis à elaboração de um questionário.

Quadro 1.5.4 - Palavras que descrevem diferentes graduações, entre 0% e 100%, segundo escala de intensidade de 5 pontos.

Escola	0%	25%	50%	75%	100%
Avaliação	Muito insatisfeito	Insatisfeito	Nem satisfeito nem insatisf.	Satisfeito	Muito satisfeito
	Muito ruim	Ruim	Nem ruim nem bom	Bom	Muito bom
	Muito infeliz	Infeliz	Nem feliz nem infeliz	Feliz	Muito feliz
Capacidade	Nada	Muito pouco	Médio	Muito	Completamente
Frequência	Nunca	Raramente	Às vezes	Repetidamente	Sempre
Impacto	Nenhum	Pequeno	Médio	Grande	Muito grande

Fonte: Likert (1932).

1.6 ELABORAÇÃO DA PLANILHA COM OS REGISTROS

Inicialmente, é preciso decidir quais instrumentos serão utilizados na coleta da informação, ou seja, elaborar uma ficha ou uma tabela onde se anotarão os dados para, posteriormente, registrá-los em computador. Em nosso caso, pode-se montar a planilha com as medidas coletadas, utilizando-se os protocolos de agilidade e ritmo, conforme os exemplos a seguir:

Durante a realização do teste, o executor deverá manter próxima a ficha de anotação dos resultados e registrar o valor de cada medida feita no pré-adolescente. Se o pesquisador elaborar uma ficha separada, para cada adolescente, aqui representado por id, deverá, em seguida, transcrever os resultados em uma planilha resumo, como nas Tabelas 1.6.1 e 1.6.2. (Recortes dos apêndices A e B respectivamente).

1.6.1 Como criar um banco de dados

No SPSS os dados podem ser tanto digitados no teclado como lidos de um arquivo externo. Os dados no SPSS são digitados diretamente na janela Editor de Dados, como numa planilha eletrônica.

Ao se criar um banco de dados, o primeiro passo é definir as variáveis, no *Variable view* (visualização das variáveis). Na coluna Nome (*Name*), criar um nome para a variável com o máximo de 8 caracteres. Em *Type* definir se é numérica, especificando o número de casos decimais (*Decimals*). Se a variável for categórica, especificar o número de caracteres que compõem o nome de cada categoria (*Width*). Na coluna *Label* (rótulo da variável), você pode digitar o nome completo da variável e será este nome que vai aparecer na janela de *output* (resultados). Em *Values* podem ser criados rótulos para as variáveis. Os valores *Missing* (valores omissos) podem ser representados por vazios na hora da digitação. *Align* determina o alinhamento da informação dentro da célula. A coluna *Measure* classifica a variável em escalar, ordinal ou nominal.

Exercício: criar um banco de dados para inserir os dados que originaram a Tabela 1.6.1. Em caso de dificuldades consulte SPSS_o_essencial (MARGOTTO, 2012).

Tabela 1.6.1 - Planilha resumo das medidas sobre o teste de lateralidade

Id	Mu	ESC	IDAD	Sex	Pes_Med	Alt_Med	Dsv T1	Dsv T2	Dsv T3	Dsv Tot	Tempo	Esq	Dir	Sb	id	Reiniciou
1	1	1	10	0	40,8	1,440	3	6	3	6	16,82	1	1	1	1	0
2	1	1	11	0	30,1	1,425	6	6	6	6	15,69	1	1	1	1	0
...																

Legenda: Id: identificador do indivíduo coletado; Mu: Município (1 = Dourados, 2 = Jardim); ESC: Escolas (Joaquim Murinho, Aurora Pedroso, Clarice Bastos, Franklin Azambuja e Etalvío Penzo); de tal forma que na planilha deve aparecer somente números. A planilha deve vir acompanhada de uma legenda especificando o que significa cada número registrado. Para um melhor entendimento vide Apêndices A e B.

Tabela 1.6.2 - Planilha resumo do teste ESS_BR (Escala de Sono)

Id	IDADE	SEXO	S1SL	S2ATV	S3SLP	S4ACP	S5DTD	S6SCA	S7SOAA	S8CPT
1	41	1	0	0	0	0	3	0	1	0
2	35	0	0	0	1	1	2	0	1	0
...										

Legenda: ESS_BR: Escala de sono de Epworth, versão em Português do Brasil; Id: identificador do indivíduo coletado; S1SL: Situação um, sentado lendo; S2ATV: Situação dois, Assistindo TV, ...

1.6.2 Crítica dos dados para detectar observações atípicas

Os erros na coleta e transcrição dos dados para as fichas e o computador traduzem-se em vieses de informação. Verificar e corrigir os erros de preenchimento que sempre se cometem na introdução (transcrição) dos dados. Para isto apresentamos três procedimentos:

i) quando as variáveis são numéricas converta todas as variáveis em escore padronizado. Procedimento para o aplicativo SPSS: Análise {Estatística descritiva < Descritiva < selecionar as variáveis e marcar a opção < salvar valores padronizados para as variables ...}. Caso o valor padronizado ultrapassar 2,5 (para amostras pequenas até 80 observações) a observação será considerada atípica. No caso de amostras maiores, este valor de 2,5 deverá ser aumentado gradativamente até 4.

ii) quando as variáveis representam medida de ordem, neste caso o gráfico Box plot (de caixas) constitui-se numa ferramenta útil.

iii) quando as variáveis são categóricas é conveniente ver a base de dados em matriz, ou seja, listando todos os valores de uma ou mais variáveis numa tabela. Este procedimento será detalhado no tópico 1.7.

A observação atípica só deve ser eliminada quando se confirmar que houve algum erro na sua obtenção.

1.7 APRESENTAÇÃO DOS DADOS DAS VARIÁVEIS EM TABELAS

Após a transcrição dos questionários em planilhas, convém organizar os dados de suas variáveis de maneira prática e racional, para o melhor entendimento do fenômeno que se está estudando.

Segundo as Normas de Apresentação Tabular do Sistema Estatístico Nacional-SEM, uma tabela deve ter número, inscrito no topo, e sua identificação deve ser feita com algarismos arábicos de forma crescente, precedidos da palavra **Tabela**.

O **Título** deve ser escrito no topo para indicar a natureza e abrangências geográfica e temporal dos dados numéricos. Estas indicações

devem ser feitas sem abreviações, por extenso, de forma clara e concisa.

O enunciado deve conter o suficiente para que sejam respondidas as seguintes questões:

- O que? (Referente ao fato)
- Onde? (Referente ao lugar)
- Quando? (Correspondente à época).

Quanto à apresentação da tabela, nela não pode haver linha divisória vertical nem horizontal e os traços horizontais devem ser apenas para isolar o cabeçalho da tabela e para finalizá-la.

Cabeçalho, deve conter palavras que indicam de forma clara e concisa o conteúdo das colunas, recomendam-se palavras escritas por extenso e sem abreviação.

Indicador de linha, inscritos nas colunas indicadoras, para indicar, complementarmente ao título, o conteúdo das linhas.

Dados numéricos são inscritos nas células para informar a quantificação de um fato específico observado.

Sinal convencional é um símbolo inscrito em uma célula, sempre que houver necessidade de se substituir um dado numérico, que deve ser feita por um dos seguintes sinais:

- Dado numérico igual a zero não resultante de arredondamento;
- .. Não se aplica dado numérico;
- ... Dado numérico não disponível;
- x Dado numérico omitido a fim de evitar a individualização da informação;
- 0 ou 0,0 etc. Dado numérico igual a zero resultante do arredondamento de um dado numérico positivo;
- 0 ou -0,0 etc. Dado numérico igual a zero resultante do arredondamento de um dado numérico negativo.

Chamada uma tabela deve ter chamada, inscrita em qualquer um de seus espaços, sempre que houver necessidade de se remeterem a alguns de seus elementos a uma nota específica.

Fonte inscrita a partir da primeira linha do seu rodapé, para identificar o responsável ou responsáveis pelos dados numéricos. Observação: no caso de publicações que contenham tabelas com dados numéricos resultantes de uma única fonte, já identificada na própria publicação, é dispensável a apresentação da fonte em cada uma das tabelas.

Nota geral, deve vir no rodapé, após especificar a fonte sempre que houver necessidade de se esclarecer o seu conteúdo geral.

Nota Específica deve constar no rodapé após a nota geral (quando esta existir), sempre que houver necessidade de se esclarecer algum elemento específico. Deve ser precedida da respectiva chamada.

Quando uma tabela contiver mais de uma nota específica, elas devem ser distribuídas de acordo com a ordem de numeração das chamadas, separando-se uma das outras por um ponto.

Ao estudar uma variável, o maior interesse do pesquisador é conhecer a distribuição dessa variável, através de suas possíveis realizações (valores). Uma maneira de se dispor um conjunto de valores de modo a se obter uma boa ideia global de sua distribuição é usando a **distribuição de frequência**. Uma medida bastante útil na interpretação de tabelas de frequências de variáveis qualitativas é a proporção de cada categoria em relação ao total, isto é, $f_i = F_i/n$. Ou, então, a porcentagem dada por $100 \times f_i$, que reduzem a frequência a um mesmo total (no caso 100).

Exemplo: Na Tabela 1.7.1 está representada a distribuição de frequência da variável qualitativa “idade do aluno”, considerando os dados de Dourados da planilha lateralidade, Apêndice A.

Tabela 1.7.1 - Frequência segundo a idade dos alunos da amostra sobre lateralidade, Dourados, 2010.

Idade ¹ [anos]	Frequência [F_i]	Proporção [f_i]	Porcentagem [%]
09	80	0,5517	55,17
10	59	0,4069	40,69
11	6	0,0414	4,14
Total	145	1,0	100

¹ Refere-se a alunos indígenas.

Refaça a tabulação usando um aplicativo (Excel), por exemplo. Abra no Excel a planilha sobre lateralidade e coloque o cursor na primeira célula, A₁₁, {vá em Inserir>Tabela dinâmica>Marque nova planilha>Ok} Aparece uma tabela modelo e ao lado uma lista de variáveis (escolha IDADE e arraste em Rótulos de linha. Depois escolha a mesma variável idade e arraste no quadrado, soma de valores. Confira os valores ao lado. Os mesmos foram transcritos na Tabela 1.7.1.

1.7.1 Classificação das séries estatísticas

Conforme critério de agrupamento, as séries estatísticas classificam-se em:

i) Série cronológica, temporal, evolutiva ou histórica.

É a série estatística em que os dados são observados segundo a época de ocorrência (anos, meses, semanas, horas e outras denominações de tempo). Elas apresentam o tempo e a característica variáveis qualitativas, enquanto que o local e o fato permanecem fixos.

Exercício: Elabore uma tabela com dados simulados para resumir o número médio de desvios da turma X, no teste de lateralidade ao

longo dos cinco primeiros anos do ensino fundamental no T1 Vel. 12.

Obs. Nas duas planilhas anexas não foram coletados dados repetidos no tempo. Portanto os dados referentes ao exercício são valores simulados e expressão apenas mais uma possibilidade de análise.

ii) Série geográfica

Os dados são observados segundo a localidade de ocorrência (municípios, zonas, inspetorias, delegacia de ensino e outros). Elas apresentam o local e a característica variáveis qualitativas, enquanto que o tempo e o fato permanecem fixos.

Exercício: Elabore uma tabela para representar o percentual de alunos por município que reiniciaram ao menos uma vez o teste de rimo. Extrair os dados da tabela do Apêndice A.

iii) Série específica

É a série estatística em que os dados são agrupados segundo a modalidade de ocorrência (fixos espaço geográfico e tempo). É, também, variável qualitativa a qual classifica os elementos de uma população de modo a separá-los em subconjuntos mutuamente exclusivos.

Exercício: Elabore uma tabela para resumir o percentual de alunos por escolas de Dourados que não reiniciaram o teste de ritmo, segundo dados apresentados na tabela do Apêndice A.

Exercícios suplementares

- a) Elabore a mesma tabela para alunos das escolas do município de Jardim.
- b) Que outras séries específicas poderiam ser organizadas em tabelas de frequência, ver dados nas tabelas dos dois anexos?
- c) Verifique quais destas variáveis são importantes e contribuem para melhorar o conhecimento dos alunos em estudo? Aponte os achados de forma resumida e faça comentários pertinentes.

iv) Distribuição de frequência para variáveis quantitativas

É a série estatística em que os dados são agrupados com suas respectivas frequências absolutas.

Exemplo 1.7.1: Frequência dados simples

Tabela 1.7.2 - Frequência por número de desvios no teste de deslocamento lateral T1 velocidade 12, dos alunos de Dourados, Apêndice A.

X_i : Número de desvios	Frequência
0	21
1	19
2	6
3	8
4	0
5	4
6	91

Exemplo 1.7.2: Frequência dados agrupados em classe.

Tabela 1.7.3 – Número de alunos por intervalo de peso dos alunos de Dourados que compõe a amostra do Apêndice A.

Faixa de peso	Frequência
20,0 – 25,2	14
25,2 – 30,4	52
30,4 – 35,6	40
35,6 – 40,8	28
40,8 – 46,0	07
46,0 – 51,2	07
51,2 – 56,4	01

Enquanto as apurações das demais séries não oferecem dificuldade, a distribuição de frequência requer característica específica de apuração. Para isto reservou-se um tópico adiante, 1.9, para o estudo desta distribuição.

1.7.2 Tabelas de dupla entrada

Refere-se à representação conjunta, na mesma tabela, dos dados de mais de uma variável categorizada, determinando a relação cruzada entre elas. Além disto, também é possível determinar as distribuições marginais e condicionais.

No exemplo sobre agilidade, pode-se extrair diversas tabelas cruzadas, como o exemplo apresentado na Tabela 1.7.4, a seguir, onde cruzaram-se a informação de habilidade de deslocamento lateral com a de sexo (gênero), dos alunos de Dourados. A habilidade de deslocamento lateral é medida de acordo com o número de vezes que o adolescente reiniciu o teste sendo classificado da seguinte forma: zero (não reiniciou o teste), disse que o aluno tem habilidade; um (reiniciou uma vez), houve falha; 2 (reiniciou duas vezes), o aluno tem pouca habilidade; três (o aluno reiniciou o teste três ou mais vezes), apresenta grande dificuldade.

Tabela 1.7.4 - Frequência e porcentagem nos testes de deslocamento lateral por categoria de sexo dos alunos, Dourados, 2010.

Em relação ao deslocamento lateral apresentou:	Sexo				Total	%
	Masculino		Feminino			
	N	%	N	%		
Habilidade	25	42,4	53	63,1	78	54,5
Falha	19	32,2	14	16,7	33	23,1
Pouca habilidade	6	10,2	15	17,9	21	14,7
Grande dificuldade	9	15,3	02	02,4	11	07,7

Nota: por motivos de arredondamento a soma dos percentuais das colunas 3 e 5 não fecham 100%

Verifica-se que 78 (54,5%) dos adolescentes não apresentaram falhas no deslocamento lateral e que 65 (45,5%) apresentaram ao menos

uma falha. Observa-se, também, que a adolescente (sexo feminino) apresentou mais habilidade de se deslocar lateralmente do que o adolescente (sexo masculino).

Exercícios:

a) Elabore a mesma tabela para alunos do município de Jardim. (Utilize um aplicativo, Excel ou SPSS ou outro)

b) Que outras variáveis poderiam ser organizadas em tabelas de dupla entrada (contingência)?

c) Verifique quais são importantes e que contribuem para melhorar o conhecimento de nossos alunos? Aponte os achados de forma resumida e faça comentários pertinentes.

1.8 GRÁFICOS

A representação gráfica das séries estatísticas tem por finalidade dar uma ideia, a mais imediata possível, dos resultados obtidos, permitindo chegar-se à conclusão sobre a evolução do fenômeno ou sobre como se relacionam os valores da série. Há várias maneiras de se representar graficamente uma série estatística. Porém, os elementos simplicidade e clareza devem ser considerados quando da elaboração de um gráfico. Apresentaremos a seguir alguns tipos de gráficos.

1.8.1 Gráfico Sequencial, Gráfico de Tendência ou Diagrama por Linha Poligonal

Objetivo: pesquisar tendências dos dados ao longo de um período, o que poderia indicar causas especiais de variação.

Na construção de um gráfico, recomenda-se: proporcionalidade (altura com três quartos a dois terços da largura), título, legenda, identificador dos eixos e escala uniforme em cada eixo.

Exemplo: usar os seguintes resultados obtidos pela turma X, na escola Y, nos quatro anos do ensino fundamental, na variável número

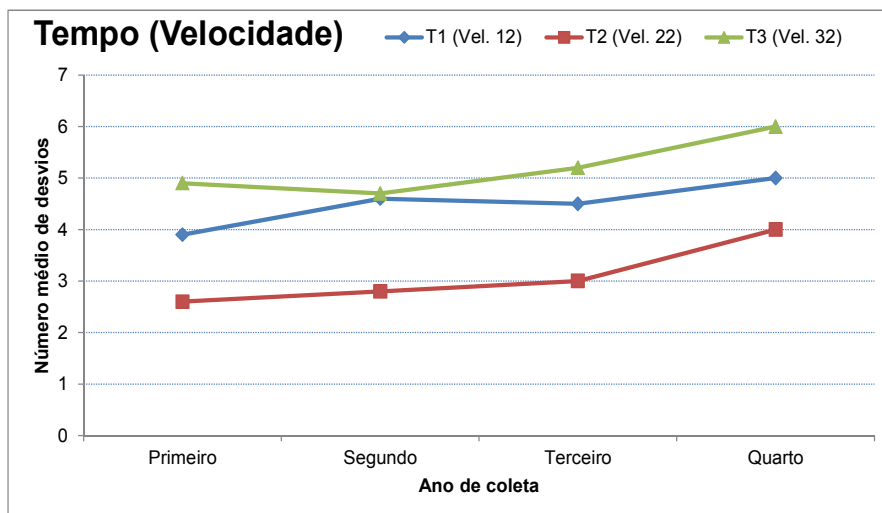
de desvios médios apresentados nas três velocidades, que tornará mais evidente, mais visual, o andamento da aprendizagem dos alunos após terem frequentado aulas de Educação Física, durante o período de avaliação.

Tabela 1.8.1 - Número médio de desvios da turma X, no teste de ritmo, ao longo de 4 anos, nas três velocidades.

Velocidade	Ano			
	1 ^o	2 ^o	3 ^o	4 ^o
T ₁ Vel. 12	3,9	4,6	4,5	5,0
T ₂ Vel. 22	2,6	2,8	3,0	4,0
T ₃ Vel. 32	4,9	4,7	5,2	6,0

Obs.: Nos dados referentes às duas planilhas anexas, não foram coletados dados repetidos no tempo. Os dados apresentados na Tabela 1.8.1 são valores simulados e expressam apenas mais uma possibilidade de análise.

Gráfico 1.8.1 - Número médio de desvios no teste de ritmo.



Fonte: tabela 1.8.1, dados simulados.

1.8.2 Gráfico de Colunas

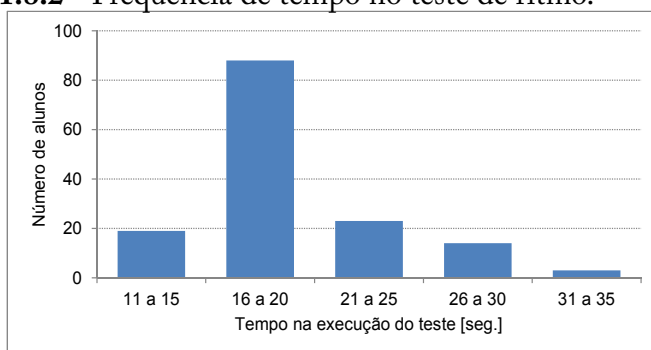
Pode ser usado para representar séries categóricas, em geral, ou séries numéricas com dados inteiros. É a representação de uma série estatística por intermédio de retângulos, dispostos em posições verticais, com espaços entre eles, **na proporção de meio espaço em branco para cada um que for usado na representação da coluna.** As demais recomendações feitas para gráfico sequencial continuam valendo para gráficos de coluna.

Exemplo: Representar através de um gráfico de colunas a distribuição do tempo que os alunos de Dourados usaram para executar o teste das três faixas (agilidade).

Tabela 1.8.2 - Frequência por intervalo de tempo para execução do teste de deslocamento, Dourados, 2010.

Tempo [seg]	Frequência	%
11 a 15	17	11,5
16 a 20	91	61,5
21 a 25	22	14,9
26 a 30	15	10,1
31 a 35	03	2,0
Total	149	100,0

Gráfico 1.8.2 - Frequência de tempo no teste de ritmo.



Fonte: tabela 1.8.2, elaborada a partir de dados do apêndice.

Observa-se que aproximadamente 60% dos estudantes demoraram entre 15 a 20 segundos para executar o teste, 11,5% o fizeram em menos de 15 segundos e 27% precisaram mais de 20 segundos para concluí-lo.

1.8.3 Gráfico em Setores

É usado para representar séries categóricas quando se pretende comparar os valores de uma categoria com a soma total. Os ângulos centrais são determinados por:

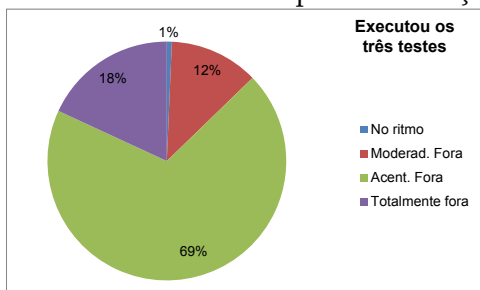
$$\theta_i = \frac{\text{Valor do setor}(i) \times 360^0}{\text{Valor total}}$$

Exemplo: represente, usando gráfico de setores, o número de estudantes que executou os três testes no ritmo, Dourados, 2010.

Tabela 1.8.3 - Resultado do teste de ritmo.

Executou	Fi	%	θ_i
No ritmo	1	0,7	2,4 ⁰
Moderad. Fora	18	12,1	43,5 ⁰
Acent. Fora	103	69,1	248,9
Totalmente fora	27	18,1	65,2 ⁰
Total	149	100	360⁰

Gráfico 1.8.3 - Percentual de alunos por classificação de acertos.



Fonte: tabela 1.8.3, elaborada a partir de dados do apêndice.

Então,

$$\theta_{\text{No ritmo}} = \frac{1 \times 360^0}{149} = 2,4^0$$

$$\theta_{\text{Mod. fora}} = \frac{18 \times 360^0}{149} = 43,5^0$$

$$\theta_{\text{Acent. fora}} = \frac{103 \times 360^0}{149} = 248,9^0$$

$$\theta_{\text{Tot. fora}} = \frac{27 \times 360^0}{149} = 65,5^0$$

Observa-se que a grande maioria, 87% dos alunos, executou o teste acentuada ou totalmente fora do ritmo.

1.9 DISTRIBUIÇÃO DE FREQUÊNCIA

Como já foi dito, a distribuição de frequência requer característica específica de apuração, para isso, o pesquisador pode lançar mão do procedimento de elaboração de tabelas de frequência que será detalhado a seguir.

1.9.1 Representação da Amostra para Variável Quantitativa Contínua

Geralmente, divide-se a amplitude dos dados brutos em um número predefinido de classes, no mínimo 5 e no máximo 12. Na literatura, existem algoritmos para se determinarem esses números. Vejamos como exemplo o peso dos alunos da amostra de Dourados. Com o uso de aplicativo, encontramos o menor e o maior valor desta variável: $X_{\text{mínimo}} = 20$ kg e $X_{\text{máximo}} = 56$ kg. A diferença entre esses dois valores representa a amplitude da variável peso dos alunos, que é igual a 36 quilos. Para determinar o intervalo de cada classe, basta dividir a amplitude pelo número de classes em que desejamos agrupar os dados da variável. No exemplo, vamos utilizar 7 classes. Esse valor pode ser calculado por meio da fórmula de Sturges ($K \cong 1 + 3,22 \times \text{Log}(n)$, onde, n = tamanho da amostra). Dessa forma, a amplitude de cada classe vale $36/7 = 5,2$ kg. Para representar o *Limite das classes* utiliza-se uma das seguintes possibilidades, a que for mais conveniente:

- ┆ Inclui o limite inferior e exclui o limite superior.
- ┆ Inclui o limite inferior e inclui o limite superior.
- ┆ Exclui o limite inferior e inclui o limite superior.
- ┆ Exclui o limite inferior e exclui o limite superior.

Isso definido, monta-se a tabela 1.9.1 da distribuição de frequência.

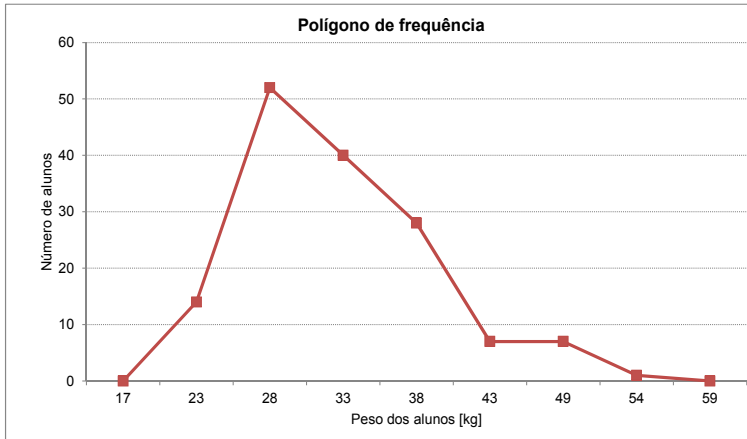
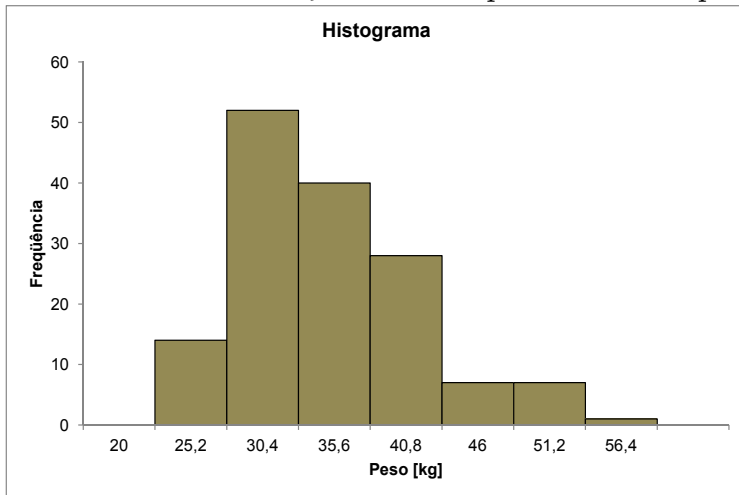
Tabela 1.9.1 - Número de alunos por intervalo de peso submetidos ao teste da lateralidade.

Classes	F_i	X_i	Fac	f_i	%
20,0 † 25,2	14	22,6	14	0,0940	9,40
25,2 † 30,4	52	27,8	66	0,3490	34,90
30,4 † 35,6	40	33,0	106	0,2685	26,85
35,6 † 40,8	28	38,2	134	0,1879	18,79
40,8 † 46,0	07	43,4	141	0,0470	4,70
46,0 † 51,2	07	48,6	148	0,0470	4,70
51,2 † 56,4	01	53,8	149	0,0067	0,67
Total	149			1,00	100

Para a contagem da frequência, recomenda-se a utilização de um aplicativo (no caso, o Excel). Procedimento: localize no Excel a planilha lateralidade, insira uma nova variável (Lim_sup_Peso), após isso, registre os 7 valores que representam o Limite Superior da Tabela 1.9.1, subtraindo 0,001 de cada um deles (essa subtração se faz necessário, pois o aplicativo Excel inclui o limite superior e exclui o inferior, e nós usaremos, o intervalo que inclui o inferior e exclui o limite superior). Salve a planilha. Em seguida, clique {Dados {Análise de dados, selecione Histograma (ok), aparece um quadro Histograma, no intervalo de entrada, selecione a variável peso, intervalo de bloco selecione a variável Lim_sup_Peso, se você marcou a primeira linha, selecione rótulo, na opção de saída marque e selecione uma célula na plan2, clique OK}} e a contagem está pronta.

O *Ponto médio das Classes* (X_i) é a soma do limite inferior da 1ª classe, com o limite inferior da 2ª classe, dividido por dois. A *Frequência absoluta acumulada* (Fac) é a soma das frequências dos valores inferiores ou iguais ao valor dado. A *Frequência relativa* (f_i) é definida como $f_i = F_i/n$.

A *representação gráfica de uma distribuição de frequência* pode ser feita através de um histograma ou de um polígono de frequência, como nos exemplos a seguir, que representa a Tabela 1.9.1

Gráfico 1.9.1 - Distribuição de alunos por intervalo de peso.**Gráfico 1.9.2 - Distribuição de alunos por intervalo de peso.**

Fonte: tabela 1.9.1, elaborada a partir de dados do apêndice.

Observa-se que existe uma assimetria positiva na distribuição dos pesos (massa corporal), sendo que, em relação aos alunos com sobrepeso, existem mais alunos com menos peso do que com sobrepeso.

Exercício: faça o mesmo procedimento para calcular a distribuição do número de alunos por intervalo de tempo gasto para executar o teste de agilidade (lateralidade).

1.9.2 Representação da Amostra para Variável Quantitativa Inteira

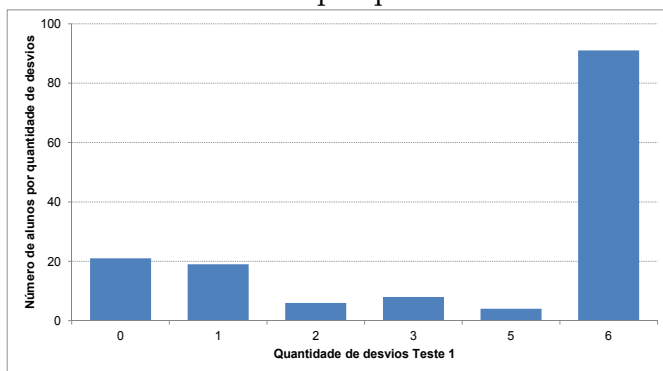
O exemplo visto na Tabela 1.9.1 refere-se à distribuição de frequência para **variável quantitativa contínua**, na qual a forma como os dados são agrupados em classe causa uma pequena perda de informação, pois considera-se que a média de todos os elementos pertencentes a uma classe como tendo o mesmo valor do ponto médio, mas isso nem sempre é verdade. Veremos, agora, um exemplo de variável quantitativa discreta, sendo que nesse caso não ocorre perda de informação, pois, o ponto X_i não é média, e sim o valor da variável.

Exemplo: distribuição de frequência do número de desvios no Teste de Ritmo, Tempo 1, dos alunos das escolas do ensino básico, Dourados, 2010.

Tabela 1.9.2 - Frequência de desvios no teste de ritmo no Tempo 1.

Desvio T1	Fi	Fac	%
0	21	21	14,1
1	19	40	12,8
2	06	46	4,0
3	08	54	05,4
5	04	58	2,7
6	91	149	61,1

Gráfico 1.9.3 - Número de alunos por quantidade de desvios no tempo 1.



Fonte: tabela 1.9.2, elaborada a partir de dados do apêndice.

Observa-se que no tempo T1, mais da metade, 61%, dos alunos apresentaram seis desvios.

1.10 MEDIDAS UNIVARIADAS

No processo investigativo, o primeiro passo é descrever o fenômeno. Para isso, a apresentação dos resultados deve ser iniciada com a análise univariada. Esta análise pode ser feita com a verificação das frequências, conforme descrito na elaboração das tabelas, o cálculo das medidas de tendência central e das medidas de dispersão para cada variável considerada no estudo. Essas medidas não poderão ser calculadas para qualquer tipo de variável, elas dependem da escala de valores de cada variável. Exemplo: não se pode calcular média de variáveis nominais. O resumo da classificação das variáveis está apresentado no Quadro 1.5.3.

1.11 MEDIDAS DE TENDÊNCIA CENTRAL PARA UMA AMOSTRA

O problema central da estatística se resume em inferir conclusões para a população, a partir de uma amostra de n indivíduos retirados aleatoriamente desta população. Antes de discutir esse aspecto, é fundamental saber informar as características da amostra. A primeira ideia que surge é o cálculo de uma medida que represente a tendência central dos dados amostrais.

As medidas de tendência central, também chamadas medidas de posição, estabelecem o valor em torno do qual os dados se distribuem. As principais medidas serão estudadas a seguir:

1.11.1 Médias

a) Média Aritmética de Dados Não-Agrupados

Definição: sejam x_1, x_2, \dots, x_n , n valores da variável X . A média aritmética simples de X representada por \bar{x} e definida por; $\bar{x} = (x_1 + x_2 + \dots + x_n) / n = (\text{soma dos valores}) / (\text{número de elementos da amostra})$.

Exemplo: em uma determinada partida de futebol foram registrados os pesos, em gramas, dos 11 jogadores, em dois momentos,

na saída do vestiário, antes do início do jogo, e na entrada do vestiário, após o fim do jogo. As diferenças entre os pesos de antes e depois da partida foram: 2.759, 3.248, 3.314, 3.101, 2.834, 3.200, 1.750, 2.900, 2.550, 3.010 e 3.500 gramas.

Calcule a perda de peso média dos jogadores.

Solução: Faça o cálculo usando um aplicativo (Excel ou calculadora).

b) Média Aritmética para Dados Agrupados

Quando os dados estiverem agrupados em uma distribuição de frequência, será usada a média aritmética, considerando-se o ponto médio das classes, dos valores x_1, x_2, \dots, x_k , ponderados pelas respectivas, frequências absolutas F_1, F_2, \dots, F_k . Isto é

$$\bar{X} = \frac{\sum_{i=1}^k x_i F_i}{\sum_{i=1}^k F_i} = \frac{x_1 F_1 + x_2 F_2 + \dots + x_k F_k}{F_1 + F_2 + \dots + F_k},$$

onde k é o número de classes.

Exercício: no exemplo Tabela 1.9.1, peso dos alunos, calcular a média aritmética.

Calcule o valor usando um aplicativo:

Excel {Dados {Análise de dados {Estatística descritiva {Intervalo de entrada (no ex. marcar a coluna dos pesos); intervalo de saída (marcar um lugar onde você quer ver os resultados de saída) e marcar resumo estatístico. Pressionar OK

$$\bar{X} = 32,5734 \text{ kg}$$

c) Média Geral

Sejam, respectivamente, $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_r$ as médias aritméticas das r séries e n_1, n_2, \dots, n_r os números de termos destas séries. A média aritmética geral formada pelos termos das r séries é dada pela fórmula:

$$\bar{\bar{X}} = \frac{n_1 \bar{X}_1 + n_2 \bar{X}_2 + \dots + n_r \bar{X}_r}{n_1 + n_2 + \dots + n_r}$$

Exemplo: A média dos pesos, para os três grupos de idade dos alunos que fizeram o teste de agilidade (lateralidade), em Dourados, está resumida na Tabela 1.11.1.

Tabela 1.11.1 - Peso médio dos estudantes por ano de idade.

Idade [anos]	Peso Médio [kg]	Número alunos
9	31,276	80
10	33,967	59
11	38,033	6

$$\bar{X} = \frac{80 \times 31,28 + 59 \times 33,97 + 6 \times 38,03}{80 + 59 + 6} = 32,65 \text{ kg}$$

Interpretação:

O peso médio geral dos alunos é 32,65 kg.

d) Média ponderada, de forma análoga a média geral, é definida como:

$$\bar{X} = \frac{p_1 X_1 + p_2 X_2 + \dots + p_r X_r}{p_1 + p_2 + \dots + p_r}$$

Exemplo: Calcular a média ponderada, do aluno Ricardo, na prova de português, no colégio Y, em 2006, MS.

Tabela 1.11.2 - Notas e pesos de quatro provas do aluno Ricardo.

Provas	Notas(X)	Pesos(P)
1	5	1
2	3	2
3	7	3
4	10	4

Fonte: dados simulados pelos autores.

Obs.: O inconveniente de \bar{x} como medida de tendência central é que a média sofre muito a influência de valores extremos. Outro aspecto a ser considerado é que só pode ser utilizada em valores quantitativos.

1.11.2 Mediana

Colocados os valores de uma variável em ordem crescente, a mediana é o elemento que ocupa a posição central da série. Isto é, deixa 50% dos valores observados à esquerda da mediana e 50% à direita. É, portanto, uma medida de classificação. Pode ser utilizada tanto em variáveis quantitativas, quanto em qualitativas ordinais.

Exemplos, para séries simples:

Seja a série 5, 7, 8, 10 e 14, a mediana será o 8, indicada por $\tilde{x} = 8$.

Seja a série 5, 7, 8, 10, 14 e 15, a mediana será o 9, indicada por $\tilde{x} = 9$.

De maneira ampla, usamos o percentil, que nos permite classificar a distribuição em 100 partes. O seu cálculo, para dados não agrupados em classes, pode ser feito usando-se a seguinte interpolação, para cálculo do percentil desejado:

$$\frac{n-1}{100} = \frac{X^0 - 1}{P}$$

Dessa forma, calcular a mediana equivale a calcular o percentil de ordem 50. Use a fórmula para calcular os valores encontrados nos exemplos acima.

Solução: na primeira série temos, $\frac{5-1}{100} = \frac{X^0 - 1}{50}$, ou, $X^0 = 3^0 = 8$

na segunda série temos, $\frac{6-1}{100} = \frac{X^0 - 1}{50}$, ou, $X^0 = 3,5^0 = 9$.

Igualmente, é possível determinar os quartis, os decis ou outra subdivisão de interesse.

Exemplo: calcule a mediana do peso dos alunos, usando um aplicativo {Excel {Dados {Análise de dados {Estatística descritiva {Intervalo de entrada (por exemplo: marcar a coluna dos pesos); intervalo de saída (marcar um lugar onde você quer ver os resultados de saída) e marcar resumo estatístico. Pressionar OK.

No relatório de saída, encontra-se, entre outros resultados, a Mediana = 31 kg.

Exercício: faça o mesmo procedimento para calcular a mediana da variável tempo, para executar os três testes.

Estabeleça uma escala padrão de 5 pontos para representar o tempo gasto pelos alunos na execução do teste das três faixas.

Procedimento: Abra no Excel a planilha Lateralidade {dados {Análise de dados {Ordem e percentil {Intervalo entrada marcar coluna do tempo e marcar uma célula em uma coluna vazia para intervalo de saída, ok aparecerá uma coluna com os valores ordenados. Agora basta escolher os valores correspondentes aos percentuais 20%, 40%, 60% e 80%, a saber, na mesma ordem, 15,82 seg., 17,5 seg., 18,765seg. e 22,59 segundos. Portanto, pode-se montar a seguinte escala padrão, em relação ao tempo necessário para um adolescente de 9 a 11 anos de idade executar o teste da lateralidade.

Tabela 1.11.3 - Escala de tempo para execução do teste da agilidade (lateralidade).

Tempo [seg.]	Menos de 15,82	15,82-17,5	17,5-18,77	18,77-22,59	Mais que 22,59
Classificação	Muito boa	Boa	Regular	Ruim	Muito ruim

1.11.3 Moda (Mo)

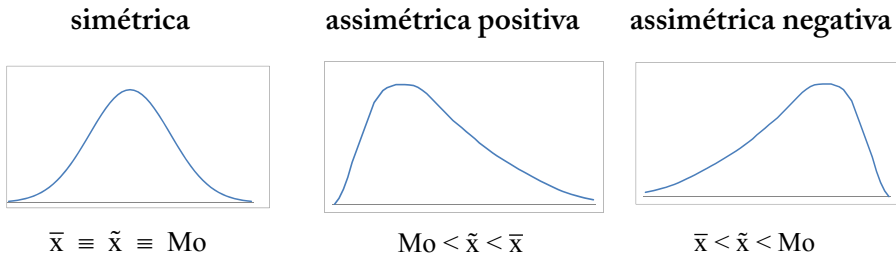
É o valor mais frequente da distribuição. Para distribuições simples, sem agrupamentos em classes, a identificação da moda é facilitada pela simples observação do elemento que apresenta maior frequência, ou seja, que mais vezes se repetiu.

Exemplo: consulte o relatório da variável peso e verifique qual o valor da moda, isto é, o peso mais comum entre os alunos $Mo = 28$ kg.

Exercício: faça o mesmo procedimento para calcular a moda da variável tempo, para executar os três testes.

Relação entre as medidas de tendência central média, mediana e moda.

Distribuição:



Outra maneira de se avaliar a assimetria da distribuição é usando o seguinte critério:

$$\bar{X} - M_0 < 0 \quad \text{assimétrica negativa}$$

$$\bar{X} - M_0 = 0 \quad \text{assimétrica}$$

$$\bar{X} - M_0 > 0 \quad \text{assimétrica positiva}$$

Exemplo: verifique qual é a distribuição dos pesos dos alunos que participaram do teste sobre lateralidade.

$$\bar{x} = 32,57 \quad > \quad \tilde{x} = 31 \quad > \quad Mo = 28$$

Portanto, a distribuição dos pesos apresenta-se como assimétrica positiva.

Exercício: verifique qual é a distribuição da variável tempo para a execução do teste. Interprete.

No Excel, Abra a planilha execução do teste depois escolha {Dados {Análise de dados {Estatística descritiva}}}. Na tabela de resultados aparece assimetria = 1,114. Como é maior do que zero, indica uma assimetria positiva.

1.12 MEDIDAS DE DISPERSÃO OU VARIABILIDADE

As medidas de variabilidade se caracterizam por medirem as diferenças entre os valores de uma distribuição. Isso significa que elas informam sobre o grau de heterogeneidade do grupo. Viu-se que as medidas de tendência central dão o valor da abscissa do ponto em torno do qual os dados se distribuem. Essas medidas serão tanto mais apropriadas para descreverem uma amostra quanto menor for a dispersão ou a variabilidade dos dados em torno da média.

Tabela 1.12.1 - Notas de quatro alunos em quatro testes diferentes.

Alunos			
A	B	C	D
5	6	10	0
5	4	5	0
5	4	5	10
5	6	0	10

É fácil observar que todos os alunos obtiveram médias iguais a 5. Pode-se, entretanto, afirmar que, em média, os alunos obtiveram uma mesma nota.

No entanto é possível estabelecer algumas afirmações sobre os dados da tabela:

- i) as notas do aluno A não variaram;
- ii) as notas do aluno B variaram menos do que as do aluno C;
- iii) as notas do aluno D variaram mais do que as dos outros.

Essas afirmações podem ser verificadas através de medidas de dispersão que serão estudadas a seguir:

Definem-se **desvios**, como sendo: $d_i = x_i - \bar{x}$.

Eles medem a dispersão entre cada valor x_i e a média aritmética \bar{x} .

Obs.: $\sum_{i=1}^n d_i = 0$.

1.12.2 Variância(σ^2)

É a média aritmética dos quadrados dos desvios. Para evitar que $\sum di = 0$ considerou-se a média dos quadrados dos desvios, $(x_i - \bar{x})^2$.

Assim:

$$\text{Var}(X) = \sigma_x^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 F_i}{n} \text{ (Populacional) ou } S_x^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 F_i}{n-1} \text{ (Amostral)}$$

Exercício: Calcular a variância do exemplo 1.12, das notas dos quatro alunos, considerando cada aluno individualmente.

Exemplo: Calcule, por meio de aplicativo, a variância da variável “peso” dos alunos que participaram do teste das três faixas.

$$S_x^2 = 45,31 \text{kg}^2$$

Exercício: Faça o mesmo procedimento para calcular a variância da variável “tempo”, para executar o teste das três faixas.

1.12.3 Desvio Padrão (σ_x ou S_x)

É a raiz quadrada da variância, isto é, $\sigma_x = \sqrt{\text{Variância}}$.

Observa-se que a variância é uma soma de quadrados, então, se a unidade da variável for, por exemplo, quilograma (kg), obtém-se como resultado kg^2 , logo, para voltar à unidade original deve-se extrair a raiz quadrada:

$$\text{No exemplo dos pesos, o } s_x = \sqrt{45,3075 \text{kg}^2} = 6,731 \text{kg}$$

Interpretação do desvio padrão

Para um conjunto de dados que apresente uma distribuição simétrica, pode-se demonstrar em estatística que, quando a amostra for grande, os intervalos relacionados abaixo:

$[\bar{x} - \sigma; \bar{x} + \sigma]$, incluem cerca de 68% das observações;

$[\bar{x} - 2\sigma; \bar{x} + 2\sigma]$, incluem cerca de 95% das observações;

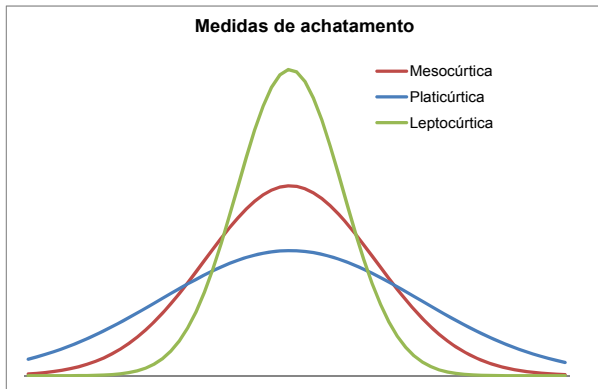
$[\bar{x} - 3\sigma; \bar{x} + 3\sigma]$, incluem cerca de 99% das observações.

Exemplo: Na variável peso dos alunos temos que o intervalo $[32,57 - 2(6,73); 32,57 + 2(6,73)] = [19,11; 46,03]$ inclui cerca de 95% do peso dos alunos.

Exercício: faça o mesmo com o tempo de execução do teste das três faixas.

1.12.4 Medidas de Achatamento ou Curtose

Como o nome indica, essas medidas procuram caracterizar a forma da distribuição quanto ao seu achatamento. O termo médio de comparação é dado pela *distribuição normal*. Assim, quanto ao achatamento, a distribuição normal é dita *mesocúrtica*. As distribuições mais achatadas que a normal são ditas *platicúrticas* e as menos achatadas são ditas *leptocúrticas*.

Gráfico 1.12.1 - Distribuição normal com diferentes achatamentos.

Fonte: Dados simulados pelos autores.

Para medir o grau de achatamento de uma distribuição utiliza-se o coeficiente C , podendo ser calculado através do uso de aplicativos, sendo que se a curva for:

- normal, $C = 0$ (a distribuição é denominada mesocúrtica);
- elevada, $C > 0$ (leptocúrtica), isto é, menos heterogênea que o normal;
- achatada, $C < 0$ (platicúrtica), isto é, mais heterogênea que o normal.

No exemplo considerado, a variável “peso”, é fácil observar através do uso de aplicativos que a curtose é igual a 0,8471. Portanto, os pesos dos alunos apresentam uma distribuição menos heterogênea que o normal, o que é esperado para a idade.

Exercício: Interprete os resultados da amostra referente aos alunos das escolas de Jardim (2011), para elaborar um relatório.

Para fechar o capítulo, reforçamos a ideia de que não é suficiente aplicar as medidas e os testes estatísticos para tirar conclusões sobre os nossos dados, pois é possível que eles não estejam corretos, ou que o desenho de estudo não seja apropriado. Desta forma, antes de se formular qualquer conclusão, é fundamental pensar em todas as possibilidades de erro nos procedimentos metodológicos envolvidos na condução da pesquisa.

Assim, na elaboração do relatório final, no capítulo sobre discussão, é importante elencar todos os problemas de forma transparente, sem tentar esconder os eventuais erros, pois, isso é grave para quem se propõe procurar a verdade. Resta-nos o consolo de que, em todos os estudos, existem limitações e cometem-se erros. Por isso, se no nosso “não houve” erros, devemos desconfiar, a não ser que a orientação tenha sido feita por alguma entidade divina.

Finalmente, a investigação deverá, sempre que possível, resultar em um trabalho publicado, pois só o que é publicado tem existência real.

Capítulo 2

Noções Sobre Probabilidade e Variáveis Aleatórias

No capítulo anterior foi mostrado como apresentar dados e como calcular medidas que descrevem características específicas desses dados. Mas o pesquisador, além de fazer tabelas e gráficos, calcular médias e desvios padrões tem sempre a pretensão de fazer inferência e testar hipóteses.

Para entender melhor essa afirmativa, imagine que o pesquisador anotou a idade, o gênero e o teste das três faixas de seus alunos. Os dados podem ser apresentados em tabelas e gráficos, podem fornecer médias e os desvios padrões.

Mas ele também gostaria de estender suas conclusões a outros alunos, além daqueles que foram submetidos ao teste, ou verificar se o resultado do teste é igual entre todos os grupos, ou se existem grupos que apresentam características diferenciadas em relação a outros.

Então, esse pesquisador gostaria de fazer inferência e testar hipóteses, além de descrever medidas.

Este livro não tem como prioridade apresentar uma fundamentação teórica mais profunda sobre o assunto. Para isso existe uma diversificada literatura que pode ser consultada: Bussab; Morettim (2011); Meyer (1982); Wonnacott; Wonnacott (1980); Fonseca; Martins (1994); e outros. A intenção aqui é apresentar minimamente os conceitos relacionados ao método científico, para que o aluno possa participar de trabalhos de pesquisas que envolvam equipes multidisciplinares e ter a compreensão mínima dos pressupostos estatísticos exigidos pelo método.

2.1 DEFINIÇÕES BÁSICAS

Todas as vezes que estudamos alguns fenômenos de observação, cumpre-nos distinguir o próprio fenômeno e o modelo matemático que melhor o explique. Para explicar os *fenômenos aleatórios* adotamos o modelo matemático probabilístico. Neste caso, o modelo utilizado será o cálculo das probabilidades.

2.1.1 Caracterização de um Experimento Aleatório não Determinístico

A fim de entender melhor essa caracterização, observemos o que há de comum nos seguintes experimentos: (Na descrição do experimento, deve-se especificar a operação ou procedimento a ser realizado, e, também, o que será observado).

- E_1 : Lançar um dado e observar o número mostrado na face superior.
- E_2 : Jogar uma moeda três vezes e observar o número de caras obtido.
- E_3 : Escolher casais com três filhos e observar a sequência de homens e mulheres obtidas.
- E_4 : Determinar a porcentagem de alunos que não consegue executar o teste no ritmo.
- E_5 : Sortear um aluno para ser “ensaiado”, quanto aos batimentos cardíacos, após ter participado de uma prova de esforço físico.
- E_6 : Medir o tempo em segundos para execução do teste de agilidade.
- E_7 : Registrar as temperaturas lidas em um termômetro, continuamente, em um período de 24 horas, em uma determinada localidade e em uma data especificada.

Dessa análise verificam-se:

- a) Cada experimento poderá ser repetido sob as mesmas condições indefinidamente;
- b) Não se conhece um valor particular do experimento *a priori*,

porém, podem-se descrever todos os possíveis resultados, as possibilidades;

- c) Quando o experimento for repetido, os resultados individuais parecerão ocorrer de uma forma aleatória. Contudo, quando o experimento for repetido um grande número de vezes, uma configuração definida ou regularidade surgirá. É esta regularidade que torna possível construir um modelo matemático preciso, com o qual se avaliará o experimento.

2.1.2 Espaço Amostral

Definição: Suponhamos que um experimento seja realizado sob certas condições fixas. Define-se espaço amostral, Ω , o conjunto de todos os possíveis resultados desse experimento.

Observação: entende-se por *resultado possível* o resultado elementar e indivisível do experimento. Todo resultado possível corresponde a um, e somente um, ponto $w \in \Omega$; e resultados distintos correspondem a pontos distintos em Ω , isto é, w não pode representar mais de um resultado.

Às vezes, o conjunto de resultados possíveis, Ω , é difícil de ser definido. Por exemplo, seja E : selecionar um habitante de Dourados/MS e medir sua altura em metros. Quais os resultados possíveis desse experimento? $\Omega = (0, \infty)$ ou $(0, 3)$ ou $[1/10, 3]$ ou ... não importa a escolha, o importante é que Ω contenha todo o resultado possível.

Exemplos 2.1: Para cada experimento enunciado acima, descrever um espaço amostral.

Solução: O espaço amostral Ω_i se refere ao experimento E_i .

$$\Omega_1$$

$$\Omega_2$$

$$\Omega_3$$

$$\Omega_4$$

$$\Omega_5$$

$$\Omega_6$$

$$\Omega_7$$

2.1.3 Eventos

Um evento, A , é um conjunto de resultados do experimento; em termos de conjuntos, é um subconjunto de Ω . Em particular Ω e \emptyset são eventos; Ω é dito o evento certo e \emptyset é dito evento impossível.

Exemplos 2.2: A_1 referir-se-á ao evento associado ao experimento E_1 .

A_1 : um número par ocorre, isto é, $A_1 = \{2, 4, 6\}$.

A_2 : {2}; isto é duas caras ocorrem.

A_3 : {HHH, HHM, HMH, MHH}; isto é, mais homens do que mulheres ocorrem.

Eventos mutuamente exclusivos: dois eventos A e B são denominados mutuamente exclusivos se eles não puderem ocorrer simultaneamente, isto é, $A \cap B = \emptyset$.

Exemplo 2.3: seja E : jogar um dado e observar o número mostrado na face de cima, então, $\Omega = \{1, 2, 3, 4, 5, 6\}$. Sejam os eventos: A = ocorrer número par e B = ocorrer número ímpar, então $A = \{2, 4, 6\}$ e $B = \{1, 3, 5\}$, ou seja, $A \cap B = \emptyset$, portanto A e B são mutuamente exclusivos.

2.1.4 Probabilidade

Se forem possíveis n repetições mutuamente exclusivas e igualmente prováveis de um experimento, e se k , dessas repetições tem determinada característica, **a probabilidade** de que ocorra um evento com essa característica é dada pela razão k/n . O resultado pode ser multiplicado por 100, para ser dado em porcentagem.

De forma simplificada, pode-se resumir assim:

$$P(A) = \frac{n^{\circ} \text{ de resultados favoráveis a } A}{n^{\circ} \text{ total de resultados em } \Omega}$$

ou

$$P(A) = \frac{NCF(A)}{NTC(\Omega)} = \frac{N^{\circ} \text{ de Casos Favoráveis a } A}{N^{\circ} \text{ Total de Casos em } \Omega}$$

Para o caso contínuo, se a área A estiver bem definida, $P(A) = \text{área } A / \text{área } \Omega$. Acontece que nem todo subconjunto do Ω tem uma área bem definida. Isto é, nem todo evento tem uma probabilidade.

Exemplo 2.4: Imagine um dado sendo lançado. Podem ocorrer os eventos: 1, 2, 3, 4, 5 ou 6.

Esses seis eventos são **mutuamente exclusivos** porque duas faces não podem ocorrer ao mesmo tempo. Se o dado não for fraudado, os seis eventos são igualmente prováveis. Nesse caso, fica fácil responder a algumas perguntas. Por exemplo, qual a probabilidade de sair número ímpar?

Dos seis eventos possíveis, três são ímpares. Então, a probabilidade de sair um número ímpar, quando se joga um dado, é igual a $\frac{3}{6} = \frac{1}{2} = 0,5 = 50\%$.

Considere outro exemplo. Uma carta será retirada ao acaso de um baralho. Qual é a probabilidade de sair um ás? Ora, um baralho tem 52 cartas, das quais quatro são ases. Então, a probabilidade de sair um ás é igual a $\frac{4}{52} = \frac{1}{13} = 0,0769 = 7,69\%$.

A probabilidade varia entre 0 e 1, ou entre 0 e 100%. Se é certa a ocorrência de determinado evento, a probabilidade desse evento é 1, ou 100%; se é impossível a ocorrência de determinado evento, a probabilidade desse evento é zero. Por exemplo, a probabilidade de ocorrer um número menor do que 8, no lançamento de um dado é 1, ou 100% (evento certo). Já, a probabilidade de ocorrer número maior que 8, é zero (evento impossível).

2.2 PROBABILIDADE CONDICIONAL

A ideia de probabilidade condicional pode ser entendida por meio de um exemplo. Imagine que um dado foi lançado. Qual é a probabilidade de ter ocorrido 5? Como o dado tem 6 faces a probabilidade de ter ocorrido a face com número 5 é $1/6 = 0,1667$ ou 16,67%.

Imagine agora que o dado foi jogado e já se sabe que ocorreu face com número ímpar. Qual é a probabilidade de ter ocorrido a face 5? Note que a resposta a esta pergunta é diferente da resposta dada à pergunta anterior. Se saiu face com número ímpar, só podem ter ocorrido os números: 1, 3 ou 5. Logo, a probabilidade de ter ocorrido 5 é $1/3 = 0,3333$ ou 33,33%.

A probabilidade de ocorrer determinado evento pode ser modificada quando se impõe uma condição. Como mostra o exemplo, a probabilidade de ocorrer 5 no jogo de um dado é 16,67%, mas, sob a condição de ter ocorrido face com número ímpar, a probabilidade de ocorrer 5 é 33,33%.

Denomina-se **probabilidade condicional** a probabilidade de um determinado evento ocorrer sob uma dada condição, que é igual ao número de eventos que possuem a característica em relação ao número de eventos do espaço reduzido pela informação (condição).

Indica-se a probabilidade condicional de ocorrer o evento A sob a condição de ter ocorrido B por $P(A | B)$, que se lê “probabilidade de A dado B”.

Como exemplo, considere a probabilidade de ocorrer um aluno com menos de três desvios no teste T1, dado que ele pertence ao sexo masculino. Essa probabilidade é condicional porque se refere à probabilidade de ocorrer um evento (ter apresentado menos de três desvios no Teste T1) sob uma dada condição (ser do sexo masculino).

2.3 EVENTOS INDEPENDENTES

Considere o seguinte experimento, lançar uma moeda e um dado e observar a sequência de resultados que aparece. O conjunto de todos os possíveis resultados do experimento está apresentado na Tabela 2.3.1, a seguir.

Tabela 2.3.1 Resultados possíveis no lançamento simultâneo de uma moeda e um dado.

Moeda\Dado	1	2	3	4	5	6
C = Cara	C; 1	C; 2	C; 3	C; 4	C; 5	C; 6
K = Koroa	K; 1	K; 2	K; 3	K; 4	K; 5	K; 6

No exemplo, podem-se extrair, entre outros, os seguintes eventos:

A: Sair cara na moeda.

B: Sair face 6 no dado.

C: Sair cara na moeda, dado que ocorreu face 6 no dado.

Neste caso, a $P(A) = 1/2$ e $P(B) = 1/18$ e $P(A | B) = 1/2$. Observe que a ocorrência de B não modificou a ocorrência de A. Neste caso dizemos que os eventos são independentes. Então, podem-se definir dois eventos independentes quando $P(A | B) = P(A)$.

Exemplificando: **Olhos claros** e **cabelos claros** são variáveis **dependentes**, porque a probabilidade de uma pessoa ter olhos claros é maior se a pessoa tiver cabelos claros. Já **olhos claros** e **idade avançada** são eventos **independentes**, porque a probabilidade de uma pessoa ter olhos claros não aumenta (ou diminui) com a idade.

2.4 TEOREMAS

Teorema do produto, se A e B são eventos independentes, então, $P(A \text{ e } B) = P(A) \times P(B)$. Se A e B não são eventos independentes $P(A \text{ e } B) = P(A) \times P(B | A)$.

Teorema da soma, suponha agora que uma caixa contenha quatro peças brancas, quatro azuis e quatro vermelhas. Retira-se uma peça da caixa ao acaso. Qual a probabilidade de ter saído peça colorida, isto é, azul ou vermelha? Ora, a probabilidade de sair peça azul é $4/12 = 1/3$ ou 33,33%, e a probabilidade de sair peça vermelha é $4/12 = 1/3 = 33,33\%$.

Então, a probabilidade de sair peça colorida, isto é, azul ou vermelha é dada pela soma: $1/3 + 1/3 = 2/3 = 66,67\%$.

Imagine agora que uma carta será retirada ao acaso de um baralho de 52 cartas. Qual é a probabilidade de sair uma carta do naipe de ouros ou um rei? A probabilidade de sair uma carta de ouro ou um rei é dada por:

$$13/52 + 4/52 - 1/52 = 16/52 = 4/13.$$

Agora fica fácil entender o teorema da soma. Se os eventos A e B não podem ocorrer ao mesmo tempo, a probabilidade de ocorrer A ou B é dada pela probabilidade de A, mais a probabilidade de B. Escreve-se:

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

Se A e B podem ocorrer ao mesmo tempo, a probabilidade de ocorrer A ou B é dada pela probabilidade de A, mais a probabilidade de B, menos a probabilidade de A e B. Escreve-se:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

Exercício:

Um casal tem dois filhos. Qual é a probabilidade de:

- o primogênito ser homem?
- os dois filhos serem homens?
- pelo menos um dos filhos ser homem?
- o segundo filho ser homem dado que o primeiro é homem?

2.5 VARIÁVEL ALEATÓRIA

Uma função X que associa a cada elemento do espaço amostral um número real $X(s)$ é denominada variável aleatória (v. a.).

Considere o exemplo de um casal ter 2 filhos. Seja X : número de filhos homens. Então,

$$S = \{(M, M); (M, H); (H, M); (H, H)\}$$

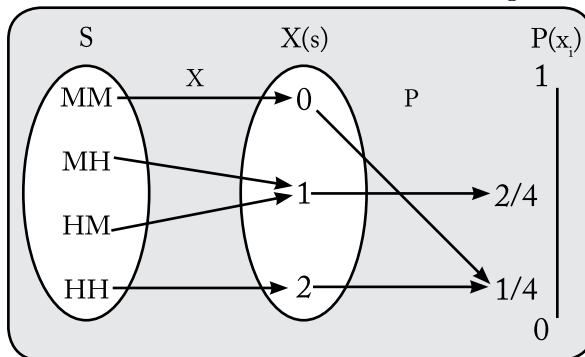
$X = 0$, corresponde ao evento (M, M) com probabilidade $1/4$;

$X = 1$, corresponde ao evento (M, H) , (H, M) com probabilidade $2/4$;

$X = 2$, corresponde ao evento (H, H) com probabilidade $1/4$.

Pode-se representar esquematicamente o resultado assim:

Figura 2.5.1 - Variável aleatória e função de probabilidade.



Fonte: dados simulados pelos autores.

2.6 DISTRIBUIÇÃO DISCRETA

Entende-se por distribuição discreta o conjunto de todos os valores que podem ser assumidos pela variável aleatória discreta, com as respectivas probabilidades. Então, o seguinte enunciado constitui um exemplo de variável aleatória discreta.

Exemplo: suponha que um professor de educação física decida fazer um teste para verificar o número de desvios com seus alunos escolhidos aleatoriamente, em uma escola de um determinado município. Seja X o número de desvios dos alunos que se submeteram ao teste no T1. Assim sendo, X é uma v. a. discreta que pode assumir os valores 0, 1, ..., 6. A probabilidade de nenhum aluno cometer desvios é 0,141; dois desvios é 0,128; ...; seis desvios é 0,611. Esta distribuição probabilística pode ser apresentada, de forma resumida, na Tabela 2.6.1.

Tabela 2.6.1 - Distribuição probabilística do número de desvios do teste no T1.

X_i	0	1	2	3	4	5	6
$P(X_i = k)$	0,141	0,128	0,040	0,054	0,000	0,027	0,611

Fonte: Dados dos autores em Apêndice A.

Note que a função $P(X = k)$ é uma função de probabilidade se a soma das probabilidades associadas a todos os valores possíveis de uma variável aleatória for sempre igual a 1 e não apresentar qualquer valor negativo. Veja no exemplo da Tabela 2.6.1 que essas condições são atendidas.

Define-se **esperança matemática** de uma v. a. discreta X como sendo a soma de todos os produtos possíveis da v. a., pela respectiva probabilidade. Isto é, $E(X) = \mu_X = \mu = \sum_i x_i p(x_i)$.

Define-se **variância** de uma v. a. como sendo $\text{Var}(X) = E(x^2) - E^2(X)$.

Exercício:

1. Calcule a $E(X)$, $\text{Var}(X)$ e interprete o resultado do exemplo da Tabela 2.6.1
2. Nas planilhas fornecidas, veja outras possibilidades de cálculos.

2.7 DISTRIBUIÇÃO BINOMIAL

A distribuição binomial é uma distribuição discreta que resulta da soma de variáveis aleatórias binárias. Exemplificando: sexo do aluno (0 = masculino, 1 = feminino), desvio no teste de deslocamento (0 = se o aluno não cometer nenhum desvio, 1 = cometer pelo menos um). Para tratar de variáveis dicotômicas, quando $n > 3$, usaremos uma fórmula que é a generalização do nosso modelo.

Por exemplo, considere a variável que representa “cometer desvio no teste da lateralidade”, para estudar a distribuição do número de vezes em que ocorreram desvios em n alunos testados, digamos $n = 5$ alunos. A primeira dificuldade que surge é calcular o espaço amostral, $\Omega = \{\text{DDDDD}, \text{DDDD}\bar{D}, \dots, \bar{\text{DDDD}}\bar{D}\}$. Imagine agora n

aumentando cada vez mais, os cálculos vão se tornando impraticáveis.

Para os problemas que envolvem $P(\text{sucesso}) = p$ e $P(\text{fracasso}) = 1 - p = q$, em n tentativas, pode-se provar que:

$$P(X=k; n, p) = C_n^k p^k q^{(n-k)}, \text{ para } k=0,1,\dots,n,$$

Sendo denominada **distribuição Binomial**.

Exemplo: Seja E: aplicar o teste da lateralidade em 5 alunos e seja $X =$ o número de alunos que apresentaram desvios no teste da lateralidade. Calcule a distribuição de probabilidade, usando a fórmula acima. Considere que a probabilidade de cometer desvios é $P(D) = 0,86$ e de não cometer desvios é $q = 0,14$.

Tabela 2.7.1 - Distribuição do número de desvios em 5 alunos escolhidos aleatoriamente.

k	$P(X = k)$	$kP(X = k)$	$k^2p(X_k)$
0	$C_5^0 0,86^0 0,14^5 = 0,0001$	0,000	0,00
1	$C_5^1 0,86^1 0,14^4 = 0,0017$	0,002	0,00
2	$C_5^2 0,86^2 0,14^3 = 0,0203$	0,041	0,08
3	$C_5^3 0,86^3 0,14^2 = 0,1247$	0,374	1,12
4	$C_5^4 0,86^4 0,14^1 = 0,3829$	1,532	6,13
5	$C_5^5 0,86^5 0,14^0 = 0,4704$	2,352	11,76
Total	1,0	4,30	19,09

$$E(X) = 4.30 = np$$

$$E(X^2) = 19,09$$

$$\text{Var}(X) = E(X^2) - E^2(X) = 19,09 - (4,3)^2 = 0,602 = npq$$

Isto é, se a distribuição for binomial, espera-se que, em média, para cada 5 alunos que se submetem ao teste da lateralidade, 4,30 deles apresentam ao menos um desvio. De outra forma, pode-se dizer que, em média, menos de um aluno (0,70) aluno, não apresenta desvio.

2.8 DISTRIBUIÇÕES DE VARIÁVEIS ALEATÓRIAS CONTÍNUAS

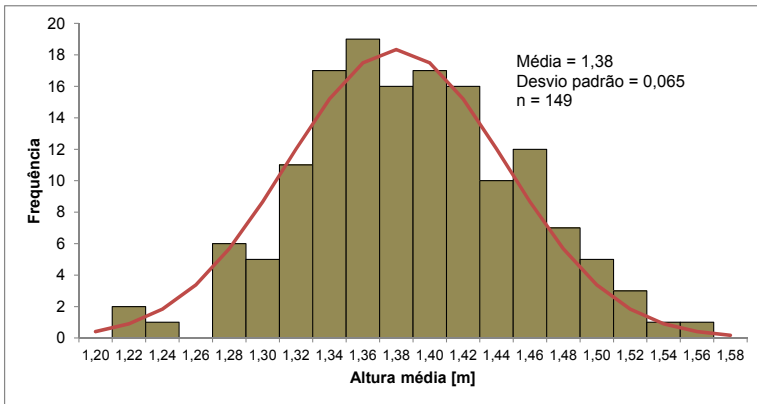
Tratam essencialmente do processo de mensuração, contrariamente à contagem ou enumeração, sendo que a principal distribuição de probabilidade, para resolver tais problemas, é a distribuição **normal ou de Gauss**.

O pesquisador estuda variáveis. O estatístico diz que estas variáveis são aleatórias porque elas têm um componente que varia ao acaso. Por exemplo, a variabilidade dos desvios no teste da lateralidade de alunos no mesmo ano de escolarização, mesma escola, mesma turma, mesmo gênero e provenientes da mesma condição social é explicada pelo acaso. Então, o número de desvios para cada aluno no teste de ritmo é uma variável aleatória.

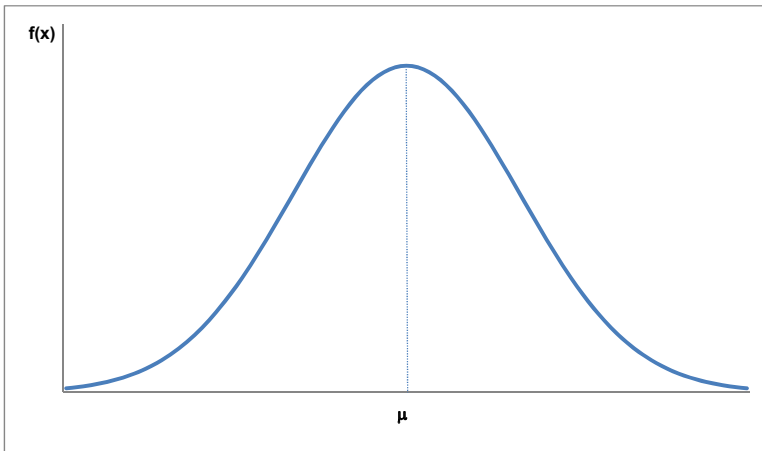
No Capítulo 1, estudou-se o comportamento de diversas variáveis, podendo ser visualizadas através do polígono de frequência ou histogramas. As variáveis estudadas e muitas outras como variáveis orgânicas, testes para medir habilidades, aptidões, erros de medidas e outros dão origem a gráficos semelhantes ao apresentado na Figura 2.8.1. Todas essas medidas são variáveis que têm distribuições que se aproximam da **distribuição normal**, apresentada na Figura 2.8.2.

A **distribuição normal** tem as seguintes características:

- a) a variável aleatória pode assumir qualquer valor real;
- b) o gráfico da distribuição normal é uma curva em forma de sino, simétrica em torno da média μ (lê-se mi), como mostra a Fig. 2.8.2;
- c) a área total sob a curva vale 1(um);
- d) como a curva é simétrica em torno da média, os valores maiores do que a média e os valores menores do que a média ocorrem com igual probabilidade;
- e) a configuração da curva é dada por dois parâmetros: a média μ e a variância σ^2 .

Figura 2.8.1 - Aproximação da distribuição normal.

Fonte: elaborada a partir de dados do apêndice.

Figura 2.8.2 - Distribuição normal.

Distribuição Normal ou Gauss, Laplace

Definição: Dizemos que a v.a. X tem distribuição normal com parâmetros μ e σ^2 , $-\infty < \mu < \infty$ e $0 < \sigma^2 < \infty$, se sua f.d.p. (função densidade de probabilidade) é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty, \text{ onde os parâmetros } \mu \text{ e } \sigma^2$$

são respectivamente sua média e sua variância.

No exemplo da Figura 2.8.1, vemos que a distribuição das alturas dos alunos que fizeram o teste da lateralidade é dada por:

$$f(x) = \frac{1}{0,065\sqrt{2(3,14)}} (1,82)^{-\frac{1}{2}\left(\frac{x-1,38}{0,65}\right)^2}, 1 < x < 2m.$$

2.9 DISTRIBUIÇÃO NORMAL REDUZIDA

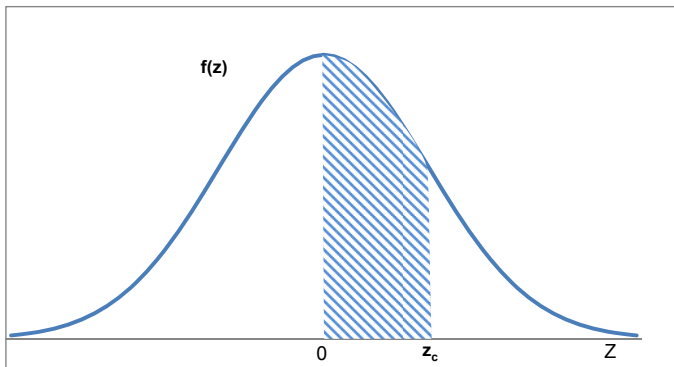
A distribuição normal reduzida – ou curva de Gauss, ou normal padrão –, é determinada através de uma transformação de variáveis, fazendo $Z = \frac{x - \mu}{\sigma}$, esta distribuição apresenta como característica

média zero e variância 1(um). A vantagem é que a probabilidade associada à distribuição normal reduzida é facilmente obtida em tabelas.

Vejamos, então, como obter probabilidades a partir da tabela de probabilidades. A Tábua I, em Anexo A, dá as probabilidades sob uma curva normal padrão que nada mais são do que as correspondentes áreas sob a curva. A Figura 2.9.1 ilustra a probabilidade fornecida pela tabela, a saber, $P(0 \leq Z \leq z_c)$. A área total sob a curva vale 1. A curva é simétrica em torno da média zero, então, a probabilidade de ocorrer valor maior do que zero é 0,5.

Mas qual seria a probabilidade de ocorrer valor entre zero e $z = 1,52$, por exemplo?

Figura 2.9.1 - Distribuição normal padrão.



A Probabilidade de ocorrer valor entre zero e $z_c = 1,52$ corresponde a área pintada na Figura 2.9.1

$P(0 \leq Z \leq 1,52) = 0,4357$ (valor fornecido diretamente pela Tábua I de valores críticos - Normais, em Anexo A).

O valor 0,4357 é localizado no corpo da tabela, pela **intersecção** da linha encabeçada pelo número 1,5 (registrado na primeira coluna da tabela) e pela coluna encabeçada pelo número 0,02 (da primeira linha da tabela). O número 1,5 compõe, com o algarismo 0,02, o número $z = 1,52$.

Observe que, devido à simetria, essa área é igual $P(-1,52 \leq Z \leq 0) = 0,4357$. Considere outro problema. Qual é a probabilidade de ocorrer valor maior do que $z = 1,52$?

Isto é: $P(Z \geq 1,52) = 0,5 - P(0 \leq Z \leq 1,52) = 0,5 - 0,4357 = 0,0643$.

Qual é a probabilidade de ocorrer valor menor do que $z = -1,52$?

Por simetria, $P(Z < -1,52) = P(Z > 1,52) = 0,0643$.

Finalmente, calcular a probabilidade de Z estar entre 0,26 e 1,52. Faça uma figura e se convença de que essa área pode ser encontrada assim:

$P(0,26 \leq Z \leq 1,52) = P(0 \leq Z \leq 1,52) - P(0 \leq Z \leq 0,26) = 0,4357 - 0,1026 = 0,3331$

2.10 COMO CALCULAR PROBABILIDADES NA DISTRIBUIÇÃO NORMAL

Como calcular probabilidades de variáveis aleatórias contínuas que seguem um padrão de distribuição normal, com média conhecida e variância conhecida? No exemplo, Figura 2.8.1, vemos que a distribuição das alturas dos alunos segue um padrão muito próximo ao da curva normal. Qual seria a probabilidade de escolher um aluno desta população e ele medir mais que 1,5 metros de altura?

Inicialmente devemos transformar a variável X : altura do estudante para a variável padrão, $Z = \frac{X - \mu}{\sigma}$.

No exemplo em discussão, $Z = \frac{x - 1,38}{0,065}$ tem distribuição normal reduzida. Se substituirmos X por $x_1 = 1,50$, temos $z_1 = (1,5 - 1,38)/0,065 = 1,85$. Então, pode-se dizer que $P(X > 1,50 \text{ m}) = P(Z > 1,85) = 0,5 - P(0 < Z < 1,85) = 0,5 - 0,4678 = 0,0322$.

Portanto, a probabilidade de escolher um aluno com mais de 1,50 metros de altura é muito pequena, 3,22%.

Exercícios: verifique outras probabilidades de interesse nas variáveis em estudo.

Capítulo 3

Inferência e Teste de Hipótese

Até o presente, referimo-nos às probabilidades de escolha de um indivíduo. Neste momento, o interesse muda e desejamos calcular probabilidade de um conjunto ou amostra de indivíduos.

3.1 DEFINIÇÕES E CONCEITOS BÁSICOS

Geralmente, estamos interessados em estudar fenômenos aleatórios. Para garantir essa premissa é necessário que a amostragem seja feita de forma probabilística, em que cada elemento da população tenha uma probabilidade conhecida de fazer parte da amostra.

Os principais tipos de amostragem probabilística são:

Amostragem aleatória simples, em que todos os “N” elementos da população têm igual probabilidade de serem escolhidos. A escolha dos “n” elementos deve ser feita através de um sorteio. Para isto é necessário que todos os elementos da população sejam enumerados, na maioria das vezes usa-se o id que já identifica o elemento na população. Depois disto, efetuar sucessivos sorteios com reposição até completar o tamanho da amostra (n). Para efetuar o sorteio usa-se a Tabua V, números aleatórios, Anexo A. **Exemplo:** retirar uma amostra simples ao acaso de 14 elementos do conjunto de alunos registrados no Apêndice A. Procedimento: inicialmente sorteia-se um número da linha, digamos 4, e um da coluna, digamos 5. Vá a Tabua V e identifique a intersecção da linha quatro com a coluna cinco, 129, que corresponde ao primeiro aluno sorteado. Depois, basta seguir esta sequência em linha ou coluna, o próximo será 169, na sequência aparece o 703 que deve ser descartado, pois a numeração dos alunos vai até o id = 180. Devemos fazer este procedimento até completar o número de elementos que compõe a amostra. No caso os idosos que foram sorteados são id = 129, 169, 100,

085, 132, 018, 110, 033, 159, 174, 132, 076, 173, 039.

Amostragem sistemática - trata-se de uma variação da amostragem simples ao acaso, muito conveniente quando a população está naturalmente ordenada, como fichas em um fichário escolar, listas de domicílios. O procedimento é muito simples, determina-se um intervalo de amostragem, que é obtido dividindo-se o total de elementos da população pelo número de elementos que pretendemos para nossa amostra (exemplo: $600/100=6$). Posteriormente, faz-se um sorteio entre 1 e 6, para determinar o ponto de início da amostragem (no exemplo, digamos que tenha sido sorteado o 2º elemento da lista), a partir daí selecionamos um elemento de 6 em 6, começando no segundo elemento da lista. Esse procedimento é feito de forma sistemática, até completar o número de elementos da amostra. **Exercício:** retire uma amostra sistemática de 12 alunos do Apêndice A.

Amostragem estratificada - neste caso a população de N unidades é, primeiramente, dividida em subpopulações de N_1, N_2, \dots, N_L unidades. Essas subpopulações não se superpõem e, juntas, abrangem a totalidade da população de tal modo que $N_1 + N_2 + \dots + N_L = N$. Essas subpopulações são denominadas de **estratos**. Depois de determinados os estratos, seleciona-se uma amostra aleatória simples em cada estrato, separadamente, de tal forma que $n_1 + n_2 + \dots + n_L = n$.

Amostragem por conglomerado quando as unidades da amostra estão agrupadas em subgrupos, consideradas de forma conglomerada. Exemplo: professores e alunos de uma instituição de ensino, diretores e operários de uma fábrica, habitantes de um bloco de apartamentos etc. Inicialmente, deve ser determinada a quantidade de escolas (se deseja estudar um fator de ensino) as quais formarão a unidade da amostra. Sortear os graus e turnos com os quais se realizará o estudo e, com a técnica aleatória simples, escolher os professores e alunos para formar as unidades da análise.

A forma de proceder para a obtenção de dados de uma amostra pode ser encontrada em Silva (1998); Bolfarine; Bussab (2005).

3.2 NOÇÃO DE ESTIMADOR

Seja (X_1, X_2, \dots, X_n) uma amostra aleatória de uma variável aleatória X , e sejam (x_1, x_2, \dots, x_n) os valores tomados pela amostra. Seja H uma função definida para os valores da amostra.

Definimos $T = H(X_1, X_2, \dots, X_n)$ como uma **estatística**, ou seja, qualquer função dos elementos da amostra.

3.2.1 Distribuições para Amostras

O problema da Inferência Estatística é fazer uma afirmação sobre parâmetros da população através da amostra. A validade de nossa afirmação será mais bem compreendida se soubermos o que acontece com a estatística T , quando retiramos todas as amostras de uma população conhecida, segundo o plano amostral adotado. Isto é, qual a distribuição de T quando (X_1, X_2, \dots, X_n) assume todos os valores possíveis. Esta distribuição é chamada de **distribuição amostral da estatística T** e desempenha papel fundamental na teoria da Inferência Estatística.

Exemplificando: Seja uma população de tamanho $N = 3$, digamos, $\{1, 2, 3\}$, então, a média da população é ; $M = \frac{1+2+3}{3} = 2$;

a variância da população é

$$\sigma_x^2 = \frac{\text{Soma de } d_i^2}{N} = \frac{(-1)^2 + (0)^2 + (+1)^2}{3} = 0,666\dots;$$

o desvio-padrão da população é

$$\sigma_x = \sqrt{0,666\dots} = 0,818\dots$$

Esses valores são denominados, respectivamente, de **parâmetros da população**, pois, no seu cálculo foi considerado todos os elementos da população.

Tabela 3.2.1 - Distribuição da amostra de algumas estatísticas obtidas da amostra de tamanho $n = 2$, retiradas da população $(1, 2, 3)$.

Amostras	Valores amostrais	Probabilidade	\bar{X}	S^2	S	X_{Max}
1	(1, 1)	1/9	1,0	0	0	1
2	(1, 2)	1/9	1,5	0,5	0,71	2
3	(1, 3)	1/9	2,0	2,0	1,41	3
4	(2, 1)	1/9	1,5	0,5	0,71	2
5	(2, 2)	1/9	2,0	0	0	2
6	(2, 3)	1/9	2,5	0,5	0,71	3
7	(3, 1)	1/9	2,0	2,0	1,41	3
8	(3, 2)	1/9	2,5	0,5	0,71	3
9	(3, 3)	1/9	3,0	0	0	3
Σ			18,0	6,0	5,66	22

Agora vamos calcular a **distribuição amostral** da média, para isso, retiram-se todas as possíveis amostras, com reposição, de tamanho $n = 2$ e calculam-se os valores da média, variância e desvio-padrão para cada amostra. Os resultados estão representados na Tabela 3.2.1

A Tabela 3.2.2 apresenta a distribuição amostral da média. Além desta, poder-se-iam descrever a da variância, a do desvio padrão e a do máximo, entre outras.

Tabela 3.2.2 - Distribuição amostral de \bar{X} .

\bar{X}_i	1,0	1,5	2,0	2,5	3,0
$P(\bar{X}_i = \bar{x})$	1/9	2/9	3/9	2/9	1/9

Em que,

$$E(\bar{X}) = \sum \bar{x}_i p(\bar{X}_i) = 2 = M, \text{ enquanto que}$$

$$\text{Var}(\bar{X}) = E(\bar{X}^2) - E^2(\bar{X}) = 4,333 - 2^2 = 0,333.$$

$$\text{É fácil ver que } \text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{0,666}{2} = \frac{\sigma_x^2}{n} = \sigma_{\bar{x}}^2.$$

Esses resultados são generalizados através dos Teoremas 3.1, 3.2 e 3.3, que são enunciados sem demonstração. Para os interessados em se aprofundar no assunto recomendamos a leitura de textos básicos de estatística como os de Meyer (1982); Bussab; Morettim (2011); Wonnacott; (1980), e outros. Esses teoremas são úteis para calcular a probabilidade de estimadores, como de média, variância, ou tamanho de amostras.

Teorema 3.1: Se X possui uma distribuição normal com média μ e desvio-padrão σ_x , então, a média amostral \bar{X} , baseada em uma amostra aleatória de tamanho n , possuirá também uma distribuição normal, com média μ e desvio-padrão $\frac{\sigma_x}{\sqrt{n}}$.

$$\text{Isto é, se } X \sim N(\mu, \sigma_x) \Rightarrow \bar{X} \sim N(\mu, \frac{\sigma_x}{\sqrt{n}}).$$

Teorema 3.2: Se X possui uma distribuição qualquer, com média μ e desvio-padrão σ_x , então, a média amostral \bar{X} , baseada numa amostra aleatória de tamanho n , terá uma distribuição que se aproxima da distribuição de uma variável normal com média μ e desvio padrão $\frac{\sigma_x}{\sqrt{n}}$ quando n for grande, maior que 30.

$$\text{Isto é, se } X \sim \forall(\mu, \sigma_x) \Rightarrow \bar{X} \sim N(\mu, \frac{\sigma_x}{\sqrt{n}}).$$

Neste caso, o teorema garante o resultado, mesmo quando não se

conhece σ_x , podendo-se usar no seu lugar o estimador S_x .

Exemplo: A professora Lourdes coletou uma amostra de $n = 149$ alunos de 4^o, 5^o e 6^o anos do ensino fundamental de Dourados/MS. O desvio padrão do peso da população é desconhecido, mas o desvio padrão da amostra é $S_x = 6,73$. Se o verdadeiro valor do peso médio da população de alunos destas séries é $\mu = 32$ kg, qual a probabilidade de que a média da amostra seja menor ou igual a 32,57 kg?

Solução: Pelo Teorema 3.2,

$$E(\bar{X}) = \mu = 32 \text{ kg e } S_{\bar{X}} = \frac{S_x}{\sqrt{n}} = \frac{6,73}{\sqrt{149}} = 0,55, \text{ então,}$$

então,

$$P(\bar{X} \leq 32,57 \mid \mu = 32 \text{ e } S_{\bar{X}} = 0,55) = P(Z \leq 1,04) = 0,5 + P(0 < Z < 1,04) = 0,5 + 0,3508 = 0,8508.$$

Onde

$$z = \frac{\bar{X} - \mu}{S_{\bar{X}}} = \frac{32,57 - 32}{0,55} = 1,04.$$

Ou seja, existem 85,08% de chances de que o peso médio de uma nova turma de 149 alunos seja inferior a 32,57kg.

3.2.2 Intervalo de Confiança para a Média

Sabe-se da distribuição normal padrão que $P(-1,96 < Z < 1,96) = 0,95$, mas $Z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}}$, então, substituindo Z pelo seu valor temos,

$$P(-1,96 < \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} < 1,96) = 0,95, \text{ ou } P(-1,96\sigma_{\bar{X}} < \bar{X} - \mu < 1,96\sigma_{\bar{X}}) = 0,95,$$

ou $P(\bar{X} - 1,96\sigma_{\bar{X}} < \mu < \bar{X} + 1,96\sigma_{\bar{X}}) = 0,95$

Em outras palavras, pode-se dizer que há 0,95 de probabilidade de que a média populacional μ esteja contida dentro do intervalo $[\bar{x} - 1,96\sigma_{\bar{x}} ; \bar{x} + 1,96\sigma_{\bar{x}}]$

No exemplo dos pesos temos que um intervalo de 95% de confiança para a média é $[32,57 - 1,96(0,55); 32,57 + 1,96(0,55)] = [31,49\text{kg}; 33,65\text{kg}]$, ou seja, podemos afirmar com 95% de confiança que o peso médio dos alunos destas séries não excede os limites de 31,49 a 33,65 kg.

Do exposto, pode-se definir **erro padrão da média**, como

$$\sigma_{\bar{X}} = \frac{\sigma_x}{\sqrt{n}}.$$

Define-se, ainda, **erro de estimativa**, como, $e = \bar{X} - \mu$.

O **tamanho da amostra** pode ser determinado, considerando que o erro de estimativa será, no máximo, o nível de significância multiplicado pelo desvio padrão, assim temos:

$$\bar{x} - \mu = 1,96\sigma_{\bar{x}}, \text{ ou } e = 1,96 \frac{\sigma_x}{\sqrt{n}} \Rightarrow n = \left(\frac{1,96\sigma_x}{e}\right)^2.$$

Exemplo, se a professora Lourdes tivesse intenção de estimar o peso médio dos alunos, com um nível de confiança de 95% e um erro máximo de 2 kg, qual deveria ser o tamanho suficiente da amostra? $n = (1,96 \times 6,73/2)^2 \cong 43,5 = 44$ alunos.

Como exercício, verifique outras situações de interesse em relação às atividades práticas usadas em tópicos anteriores.

Teorema 3.3: Se X possui uma distribuição qualquer, com média μ e desvio-padrão desconhecido, então, a média amostral \bar{X} , baseada em uma amostra aleatória de tamanho n , terá uma distribuição que se aproxima da distribuição t de Student, com média μ e desvio padrão $\frac{S_X}{\sqrt{n}}$, quando n for pequeno, menor do que 30.

Isto é, $X \sim \forall(\mu, \sigma_X \text{ desconhecido}) \Rightarrow \bar{X} \sim t(\mu, \frac{S_X}{\sqrt{n}})$.

Define-se a variável t por $t = \frac{\bar{X} - \mu}{S_X} \sqrt{n}$. Os valores de t , à semelhança de Z , encontram-se na Tábua II, no Anexo A. Apresenta-se na Tabela 3.2.3 o resumo de algumas fórmulas úteis, para estimavas de parâmetros.

Tabela 3.2.3 - Resumo de estimadores.

A. Estimativa de médias	População Infinita	População Finita*
Pontual	\bar{X}	\bar{X}
Intervalar; σ_x conhecido	$\bar{X} \pm Z_c \frac{\sigma_x}{\sqrt{n}}$	$\bar{X} \pm Z_c \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$
σ_x desconhecido	$\bar{X} \pm t_c \frac{S_X}{\sqrt{n}}$	$\bar{X} \pm t_c \frac{S_X}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$
Tamanho da amostra σ_x conhecido	$n = Z_c^2 \frac{\sigma_x^2}{e^2}$	$n = \frac{Z_c^2 \sigma_x^2 N}{Z_c^2 \sigma_x^2 + e^2(N-1)}$
σ_x desconhecido	$n = t_c^2 \frac{S_X^2}{e^2}$	$n = \frac{t_c^2 S_X^2 N}{t_c^2 S_X^2 + e^2(N-1)}$
Erro σ_x conhecido	$e = Z_c \frac{\sigma_x}{\sqrt{n}}$	$e = Z_c \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$
σ_x desconhecido	$e = t_c \frac{S_X}{\sqrt{n}}$	$e = t_c \frac{S_X}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$
B. Estimativa de proporções		
Pontual	$\hat{p} = \frac{S_n}{n}$	$\hat{p} = \frac{S_n}{n}$
Intervalar	$\hat{p} \mp Z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\hat{p} \mp Z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \sqrt{\frac{N-n}{N-1}}$
Tamanho da amostra	$n = Z^2 \left(\frac{\hat{p}(1-\hat{p})}{e^2} \right)$	$n = \frac{Z^2 \hat{p}(1-\hat{p})N}{(N-1)e^2 + Z^2 \hat{p}(1-\hat{p})}$
Erro	$e = Z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$e = Z_c \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \sqrt{\frac{N-n}{N-1}}$

* Quando o tamanho da amostra é mais que 5% do tamanho da população

3.3 REGRA DE DECISÃO E PROBABILIDADE DE ERRO TIPO I E TIPO II

Neste ponto, trataremos do seguinte problema. Feita determinada afirmação sobre a população, usualmente, sobre um parâmetro desta, desejamos saber se os resultados de uma amostra contrariam ou não tal afirmação. O objetivo do teste estatístico de hipótese é fornecer ferramentas que nos permitam validar ou refutar uma hipótese (estatística) através dos resultados da amostra.

Entende-se por **hipótese estatística** qualquer afirmação que se faça sobre um parâmetro desconhecido θ .

Baseados em uma amostra da população, vamos estabelecer uma regra de decisão, segundo a qual rejeitamos ou aceitamos a hipótese proposta. Tal regra de decisão é chamada de teste.

Exemplo: Um professor aplica um teste envolvendo dez questões, do tipo certo ou errado. Ele quer testar a hipótese “o estudante está adivinhando”.

Então, queremos pôr à prova a hipótese $H: p = \frac{1}{2}$ (hipótese que o estudante está adivinhando).

Mas para poder testar esta hipótese, o professor deve adotar uma **regra de decisão**: “se oito ou mais respostas estão corretas, o estudante não está adivinhando, enquanto que se menos de oito questões estiverem corretas, o estudante está adivinhando”.

Indicamos por X = o número de respostas certas;

S = o espaço amostral dos valores de X .

No exemplo, $S = \{0, 1, 2, \dots, 10\}$.

A hipótese H será rejeitada se observarmos os valores $x = 8$, $x = 9$ ou $x = 10$. Esses valores constituem a chamada **região crítica ou região de rejeição do teste**, e indicamos por S_0 .

No exemplo, $S_0 = \{8, 9, 10\}$, a região de aceitação do teste será $S_1 = S - S_0$.

No exemplo, $S_1 = \{0, 1, 2, 3, 4, 5, 6, 7\}$.

O professor sabe, no entanto, que é possível que o estudante esteja adivinhando e, ainda assim, ele acerte oito ou mais questões. Isto é, H é verdadeira, mas será rejeitada. O professor quer, certamente, que essa probabilidade seja pequena. É fácil calculá-la.

No exemplo tem-se:

$$P\left(X \geq 8 : p = \frac{1}{2}\right) = p(X=8) + p(X=9) + p(X=10) = \\ C_{10}^8 \left(\frac{1}{2}\right)^{10} + C_{10}^9 \left(\frac{1}{2}\right)^{10} + C_{10}^{10} \left(\frac{1}{2}\right)^{10} = \frac{56}{1024} = \frac{7}{128} = 0,054.$$

O significado da resposta é o seguinte: se o teste pudesse ser aplicado 128 vezes, o professor esperaria rejeitar H (o aluno está adivinhando) 7 vezes. A probabilidade $\frac{7}{128} = 0,054$ é **chamado nível de significância** e o indicamos por α ou **erro tipo I**.

Hipótese alternativa: suponhamos, agora, que o aluno acertou apenas seis questões. Então não há razão para rejeitar H e diríamos que o aluno está adivinhando. Mas é possível que o aluno não esteja adivinhando e acertou apenas seis questões. Portanto, há um outro erro que está envolvido neste processo decisório; **aceitar uma hipótese H , sendo ela falsa**. Assim, podemos pensar o problema em termos de testar:

$$H_0 : p = 0,5$$

$$H_1 : p = 0,8$$

No exemplo, vamos calcular a probabilidade β , de aceitar H_0 , quando H_1 é verdadeira.

$$\beta = p(X=0 \text{ ou } X=1 \text{ ou } \dots \text{ ou } X=7 : p=0,8) = C_{10}^0 (0,8)^0 (0,2)^{10} + \dots \\ + C_{10}^7 (0,8)^7 (0,2)^3 = 0,322.$$

Este erro, **aceitar uma hipótese falsa, é denominado erro tipo II** e β é a probabilidade de erro tipo II.

Trocando as regiões críticas, α e β variam conforme valores apresentados, Tabela 3.3.1.

Tabela 3.3.1 - Valores de α e β para diferentes regiões críticas, com n fixo igual a 10.

Regiões Críticas	α	β
{7, 8, 9, 10}	0,170	0,121
{8, 9, 10}	0,054	0,322
{9, 10}	0,010	0,624

Para n fixo a priori, não é possível tomar α e β à vontade, gostaríamos que α e β fossem pequenos, mas devido ao fato acima, costuma-se fixar um valor para α normalmente pequeno. A justificativa é que fornece o erro tipo I mais grave. Obs.: No momento de planejar o experimento para testar uma hipótese, o tamanho da amostra deve levar em consideração $\alpha = 0,05$ e $\beta = 0,20$.

Quadro 3.3.2 - Resumo dos erros (estado) e as decisões.

ESTADO	DECISÃO	
	Aceitar H_0	Rejeitar H_0
Se, H_0 for verdadeira	Decisão correta. Probabilidade = $1 - \alpha$ = nível de confiança	Erro tipo I. Probabilidade = α = nível de signif. do teste
Se, H_0 for falsa	Erro tipo II. Probabilidade = β	Decisão correta. Probabilidade = $1 - \beta$ = poder do teste

Normalmente, o erro tipo I deve ser considerado como o mais grave. A justificativa pode ser percebida no exemplo a seguir:

Exemplo. O problema que se propõe é o de escolher uma das seguintes hipóteses como hipótese nula:

- H_i : a vacina tem efeito positivo;
- H_j : a vacina é inócua.

Se, $H_0 = H_1$, o erro tipo I consiste em considerar a vacina inócua, quando ela é eficiente. Isto é, a vacina é eficiente, mas a consideramos inócua.

Se, $H_0 = H_1$, o erro tipo I consiste em considerar a vacina eficiente, quando ela é inócua. Isto é, a vacina é inócua, mas a consideramos eficiente.

Logo, tomamos H_0 como hipótese nula, pois o erro tipo I decorrente nos parece mais grave.

3.3.1 Passos para Construção de um Teste de Hipóteses

Discutiu-se no item 3.3 os conceitos teóricos. Apresenta-se a seguir uma seqüência que pode ser usada sistematicamente para qualquer teste de hipóteses.

Primeiro: fixe qual a hipótese H_0 a ser testada e defina qual a hipótese alternativa H_1 .

Segundo: use a teoria estatística e as informações disponíveis para decidir qual estatística (estimador) será usada para julgar a hipótese H_0 . Não se esqueça de levantar as propriedades dessa estatística.

Terceiro: fixe a probabilidade de cometer um erro de primeira espécie, e use este valor para *construir a Região Crítica (RC)*. Valores estes, normalmente, encontrados em tabelas de acordo com a distribuição usada.

Quarto: use as informações fornecidas pela amostra para calcular o valor da estatística que definirá a decisão.

Quinto: se o valor da estatística, calculado com os valores da amostra, não pertencer a região crítica, aceite H_0 , caso contrário, rejeite.

Exemplo: A professora Lourdes disse que seus alunos devem demorar, em média, 18,5 segundos para executar o teste das três faixas. Com base na amostra dos alunos das escolas de Dourados, Apêndice A, verifique se a afirmativa da professora é verdadeira?

Solução:

Primeiro: seja X : o tempo que cada aluno demorou para executar o teste. Através dos dados da amostra, Apêndice A, calculou-se $\bar{X} = 19,16$; $S_x^2 = 17,876$.

As hipóteses que nos interessam são:

$$H_0: \mu = 18,5 \text{ seg}$$

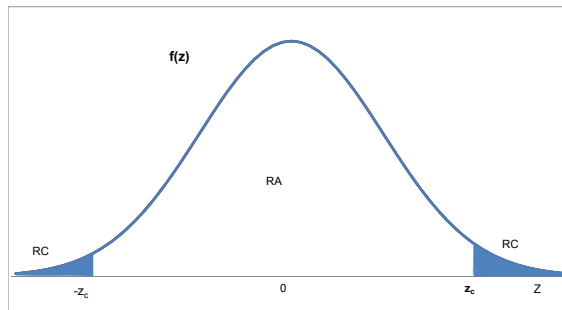
$H_1: \mu \neq 18,5 \text{ seg}$, pois o tempo médio pode diferir para mais ou para menos da média.

Segundo: como $n > 30$, o Teorema 3.2 nos assegura que podemos usar a variância da amostra, $s^2 = 17,88$, como estimativa da variância da população. Assim, qualquer que seja a média μ , a média amostral dos 148 estudantes terá a distribuição

$$\bar{X} \sim N\left(\mu, \frac{17,88}{148}\right) \text{ ou } \bar{X} \sim N(\mu, 0,1208), \text{ onde } \sigma_{\bar{X}} = \sqrt{0,1208} = 0,34756.$$

Terceiro: vamos fixar $\alpha = 5\%$; pela hipótese alternativa, vê-se que a hipótese H_0 deve ser rejeitada quando \bar{X} for muito pequeno ou muito grande (teste bicaudal). Assim, nossa região crítica será como a da figura:

Gráfico 3.3.1 - Região crítica da distribuição normal.



Fonte: elaborado pelos autores.

Do Anexo A, Tábua I, da curva normal, obtêm-se os valores críticos para z , isto é, $-z_c = -1,96$ e $z_c = 1,96$.

Logo, a região crítica ou rejeição do teste é dada por $RC = \{z_c \in \mathbb{R} | z_c < -1,96 \text{ ou } z_c > 1,96\}$.

Quarto: com as informações observadas da amostra, calculamos $Z_{\text{observado}}$

$$Z_0 = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{19,16 - 18,5}{0,34756} = 1,90$$

Quinto: como $Z_0 \notin RC$, a nossa conclusão será não rejeitar H_0 . Isto é, essa diferença é atribuída ao acaso e a afirmação da professora Lourdes não está errada.

Na literatura é comum o uso do termo P-value, que significa rejeitar H_0 com um nível de significância mínimo de 5%. Isto é, a probabilidade de rejeitar H_0 sendo ela verdadeira é menor que 0,05, ou a Probabilidade ($F_{\text{observado}} \geq F_{\text{crítico}} < 0,05$).

No exemplo $p(F_0 \geq \pm 1,90) = 0,0574$ (p-value), que é maior que 0,05, portanto, não rejeitamos H_0 .

3.4 ESTIMADORES OU ESTATÍSTICAS USADOS PARA JULGAR HIPÓTESES

Teste para a média

Admitindo-se que não conhecemos a variância populacional, a variável de teste será o t.

$$t_0 = \frac{\bar{X} - \mu_0}{s_x} \sqrt{n}, \text{ onde, } \bar{X} = \text{média amostral; } \mu_0 = \text{valor da hipótese nula;}$$

s_x = desvio-padrão da amostra e n = tamanho da amostra.

Teste de uma amostra para a proporção

Para amostras de mais de 20 observações, usa-se a distribuição normal.

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}, \text{ onde } \hat{p} = \text{valor estimado com os valores da amostra;}$$

p_0 = valor da hipótese nula; n = tamanho da amostra. Sendo $Z_0 \sim N(0,1)$.

Para $n \leq 20$, deve-se usar a distribuição Binomial.

Ao se fazer a amostragem em uma população finita (e $n/N > 5\%$), deve-se usar o fator de correção finita

$$\sqrt{\frac{N-n}{N-1}}, \text{ onde, } \sigma_p = \sqrt{\frac{p_0(1-p_0)}{n}} \sqrt{\frac{N-n}{N-1}}, \text{ então,}$$

$$Z_0 = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)(N-n)}{n(N-1)}}}.$$

Teste de duas amostras para proporções

Seja $H_0: p_1 = p_2$, em que,

$$\hat{p} = \frac{x_1 + x_2}{n_1 + n_2} \text{ e } \sigma_{\hat{p}} = \sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)},$$

então

$$Z_0 = \frac{(x_1/n_1) - (x_2/n_2)}{\sqrt{\hat{p}(1-\hat{p})(1/n_1 + 1/n_2)}}, \text{ sendo,}$$

x_1 = o número de sucessos na amostra 1; x_2 = o número de sucessos na amostra 2;

n_1 = o número de observações na amostra 1; n_2 = ao número de observações na amostra 2.

Teste de k amostras para proporções (χ^2 : distribuição quiquadrado)

A finalidade é avaliar a alegação de que k amostras independentes provenham de populações que contenham a mesma proporção de determinado item.

Seja H_0 : As proporções populacionais são todas iguais. A estatística é,

$$\chi^2_{\text{Cal}} = \sum_{i=1}^r \left[\frac{(O_i - E_i)^2}{E_i} \right] \text{ onde, } O_i = \text{frequência observada; } E_i = \text{frequência}$$

esperada e r = número de linhas

Tem uma distribuição,

$\chi_{tabel}^2 = \chi_{(r-1)(k-1)}^2$ onde r = número de linhas; k = número de colunas.

Teste para igualdade ou diferença entre duas médias

Variâncias conhecidas

Seja uma amostra X_1, X_2, \dots, X_n da população $N(\mu_1, \sigma_1^2)$ e uma amostra Y_1, Y_2, \dots, Y_n da população $N(\mu_2, \sigma_2^2)$, independentes.

Então,

$$Z_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2 / n_1 + \sigma_2^2 / n_2}}, \text{ quando conhecemos as variâncias populacionais } \sigma_1^2 \text{ e } \sigma_2^2.$$

Com $Z_0 \sim N(0, 1)$

Variâncias iguais e desconhecidas

Quando σ^2 da população é desconhecido, admitimos $\sigma_1^2 = \sigma_2^2 = \sigma^2$ e amostras independentes.

$$t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_c \sqrt{\frac{n_1 + n_2}{n_1 n_2}}}, \text{ onde}$$

$$S_c = \sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}, \text{ tem uma distribuição t de Student,}$$

com $n_1 + n_2 - 2$ graus de liberdade.

Variâncias desiguais e desconhecidas

$$t_0 = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}, \text{ tem uma distribuição t de Student,}$$

$$\text{com, } \nu = \frac{(S_1^2 / n_1 + S_2^2 / n_2)^2}{\frac{(S_1^2 / n_1)^2}{n_1 - 1} + \frac{(S_2^2 / n_2)^2}{n_2 - 1}}, \text{ g.1.}$$

Teste para a variância

$$\chi_0^2 = \frac{(n-1)S^2}{\sigma_0^2}, \text{ onde,}$$

n = tamanho da amostra; S^2 = variância amostral; σ_0^2 = valor da hipótese nula; tem uma distribuição χ_{n-1}^2 .

Teste para a igualdade de duas variâncias

Seja uma amostra X_1, X_2, \dots, X_n da população $N(\mu_1, \sigma_1^2)$ e uma amostra Y_1, Y_2, \dots, Y_m da população $N(\mu_2, \sigma_2^2)$, independentes.

Seja $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma$, então,

$$F_0 = \frac{S_1^2}{S_2^2}, \text{ tem uma distribuição F}(n-1, m-1).$$

Capítulo 4

Formulação de Hipótese Estatística, Teste Adequado, Apresentação do Resultado

Neste capítulo apresentaremos complementarmente outros exemplos de hipóteses que podem ser formuladas sobre os testes de lateralidade e ritmo, cujos resultados da pesquisa estão apresentados na planilha lateralidade (Apêndice A). Será apresentado o teste adequado com respectivo procedimento e justificativa. Na sequência, os resultados serão apresentados de forma abreviada. A Tabela 1.5.3 deverá servir de referencial para a escolha do tipo de teste apropriado. O embasamento teórico para alguns dos testes utilizados neste capítulo foge do escopo deste livro. Os interessados poderão consultar literatura específica sobre análise de variância, correlação, regressão linear, estatística não paramétrica, gráficos box-plot, e outros.

4.1 COMPARAR DUAS MÉDIAS COM PRESSUPOSTO DE NORMALIDADE

Hipótese: Em Dourados, os meninos demoram menos tempo em relação à execução do teste das três faixas do que as meninas.

Experimento: Anotar o tempo de execução do teste da lateralidade em três faixas de velocidade; Teste: t de Student.

Procedimento: Abra no SPSS a Planilha Lateralidade. {Análise> Comparar médias> Teste t para amostras independentes> Indicar, como variável de teste, o tempo de execução do teste das três faixas e, como variável categórica, o sexo, com categorias: 0 para feminino, 1 para masculino> Clicar OK}.

Justificativa: As variáveis envolvidas são tempo de execução (dependente, quantitativa) e gênero (independente, categórica, com apenas dois grupos: Masculino, Feminino). Foi testada a normalidade da variável tempo de execução e esta se apresentou normal. Caso isso não tivesse acontecido, seria necessário aplicar a prova U de Mann-Whitney (prova não paramétrica). Também foi testada a igualdade de variâncias, através do teste de Levene, cujo resultado aponta para a igualdade de variâncias. No relatório de saída do SPSS são fornecidos os resultados dos testes, tanto para variâncias iguais, como para variâncias diferentes. Deve-se ler o resultado na linha apropriada para a interpretação correta do p -valor. O pressuposto de independência entre os grupos é automático, pois são grupos vindos de populações mutuamente excludentes: meninos e meninas.

Tabela 4.1.1 - Distribuição das médias e demais medidas do tempo de execução do teste em relação ao gênero.

Gênero	Nº de sujeitos	Média	Desvio padrão	t de Student	p-valor
Feminino	87	19,78	4,41	t(146) = 2,146	0,033
Masculino	61	18,28	3,82		

Caso não seja interesse do pesquisador elaborar uma tabela com tantos dados, é possível criar a tabela apenas com os dados descritivos (n , média, desvio padrão) e, no texto, apresentar o resultado da prova tempo de execução do teste das três faixas.

Apresentação do resultado da prova tempo de execução do teste das três faixas no texto:

$t_{(\text{graus de liberdade})}$ = valor de t , p = valor da significância da amostra.

No exemplo acima: $t_{(146)} = 2,146$, $p = 0,033$.

Exercício: acrescente na Tabela 4.1.1 o resultado das variáveis peso e altura dicotomizado com 0: menor igual a média e 1: maior que a média, processe no SPSS o teste com essas duas variáveis indicadoras e veja na sequência como fazer a apresentação dos resultados.

Observa-se que o tempo de execução do teste das três faixas não é igual entre os sexos, sendo que, o tempo médio para execução diferiu significativamente ($t_{(146)} = 2,146$; $p = 0,033$), entre adolescentes masculinos e femininos, pelo teste *t* de Student. O tempo médio de execução do teste foi significativamente, ($t_{(146)} = 2,437$; $p = 0,016$), diferente para a variável peso, onde, os mais “leves” demoraram mais tempo, em média, para executar o teste. Os adolescentes de menor estatura demoraram mais tempo, em média, para executar o teste das três faixas, porém, a diferença não foi estatisticamente significativa ($t_{(146)} = 1,73$; $p = 0,086$).

4.2 COMPARAR DUAS MÉDIAS SEM PRESSUPOSTO DE NORMALIDADE

Hipótese: Andar no ritmo com três velocidades diferentes, em Dourados, em média, os meninos cometeram menos erros em comparação com as meninas.

Experimento: executar o teste de agilidade em três faixas de velocidade e anotar o número de desvios cometidos pelo aluno; **Teste** *U* de Mann-Whitney, *in* Siegel (1981).

Procedimento: Abra no SPSS a Planilha Ritmo. {Análise> Testes não paramétricos> 2 amostras independentes> Teste *U* de Mann-Whitney> Indicar como variável de teste o número médio de desvios nos testes com três ritmos e a variável categórica sexo com categorias: 0 para feminino, 1 para masculino> clicar OK}.

Justificativa: As variáveis envolvidas são números médios de desvios nos testes com três ritmos (dependente, quantitativa) e gênero (independente, categórica, com apenas dois grupos: Masculino, Feminino). Foi testada a normalidade da variável dependente e esta não seguiu a distribuição normal. Como isso não aconteceu, foi necessário aplicar a prova *U* de Mann-Whitney (prova não paramétrica). Também foi testada a igualdade de variâncias, através do teste de Levene, cujo resultado aponta para a não igualdade de variâncias. O teste foi executado no aplicativo SPSS e o resultado foi gerado automaticamente, tanto para variâncias iguais, como para variâncias

diferentes. É só ler os resultados na linha apropriada. O pressuposto de independência entre os grupos é automático, pois são grupos vindos de populações mutuamente excludentes: meninos e meninas.

Tabela 4.2.1 - Distribuição das médias do número de erros e demais medidas desta variável em relação ao gênero, peso e altura.

Variáveis	Adolescente	N	Média erros	Desvio padrão	U de Mann-Whitney	p-valor
Sexo	Feminino	87	4,28	1,29	Z = -3,362	0,001
	Masculino	62	3,39	1,68		
Peso	Abaixo da média					
	Acima da média					
Altura	Abaixo da média					
	Acima da média					

Apresentação do resultado da prova Teste da lateralidade em três faixas de velocidade: Z = valor do teste, p = valor da significância da amostra, conforme exemplo da Tabela 4.2.1: $Z = -3,362$, $p = 0,001$ e assim por diante ver os demais casos.

Exercício: acrescente na Tabela 4.2.1 o resultado das variáveis peso e altura dicotomizado com, 0: menor igual à média e 1: maior que a média, processe no SPSS o teste com essas duas variáveis indicadoras e veja na sequência como fazer a apresentação dos resultados.

Observa-se que o número médio de erros na execução do teste das três faixas não é igual entre os sexos, sendo que, para os alunos, o número médio de erros para execução foi significativamente ($Z = -3,362$; $p = 0,001$) menor do que o apresentado pelas alunas, no teste U de Mann-Whitney. O número médio de erros na execução do teste não diferiu significativamente para a variável peso, ($Z = -0,251$; $p = 0,801$), ou seja, apesar das diferenças individuais de peso, os erros estiveram na média, indistintamente. No mesmo teste, também não se observou diferença significativa de erros, associados à estatura dos avaliados ($Z = -0,624$; $p = 0,533$).

4.3 ASSOCIAÇÃO ENTRE DUAS VARIÁVEIS CATEGÓRICAS

Hipótese: Executar o teste de ritmo no tempo, depende do sexo, peso e altura

Experimento: Executar o teste de ritmo e observar se o estudante andou no ritmo; Teste: Qui-quadrado (χ^2) de associação entre duas variáveis, (DAWSON; TRAPP, 2003).

Procedimento: Abra no SPSS a Planilha Ritmo e Lateralidade. {Análise> Estatística descritiva> Tabela cruzada> Indicar na linha a variável sexo e na coluna a variável executou o teste na velocidade 1> selecionar estatística e marcar Qui-quadrado. Pressione continue e clique OK}. O procedimento deverá ser repetido para as variáveis peso e altura.

Justificativa: a variável andar no ritmo na execução dos testes com três velocidades é categórica e não atende aos pré-requisitos necessários à execução de um teste paramétrico.

Tabela 4.3.1 - Frequência no teste T1 de ritmo, por sexo, peso e altura dos participantes e significância no teste.

Variáveis	Adolescente	Teste de ritmo		$\chi^2(p)$	p-valor
		No ritmo ¹	Fora do ritmo		
Sexo	Feminino	21	66	4,443	0,035
	Masculino	25	37		
Peso	Abaixo da média	44	33	0,444	0,505
	Acima da média	45	27		
Altura	Abaixo da média	46	31	0,000	0,998
	Acima da média	43	29		

1 Andar no ritmo significa não cometer nenhum desvio no teste de ritmo.

Exercício: refazer a Tabela 4.3.1 com o resultado das variáveis Teste de Ritmo no tempo T2 (vel. 22) e T3(vel. 32).

Apresentação dos resultados

Em relação ao teste 1, de ritmo, constatou-se, conforme a tabela 4.3.1, um número alto de adolescentes, 103 (69,1%), que não conseguiram executar o teste no ritmo, sendo 66 (75,9%) adolescentes do sexo feminino e 37 (59,7%) do sexo masculino. Essa diferença foi estatisticamente significativa ($\chi^2(1) = 4,443$; $p = 0,035$), ou seja, os indivíduos pertencentes às duas amostras diferiram em relação ao teste 1 de ritmo, sendo o sexo feminino o que apresentou maior dificuldade na sua execução. No teste 2, o número de adolescentes que não conseguiu executar o teste no ritmo foi de 60 (40,3%), sendo que, destes, 18 (29,0%) são do sexo masculino e 42 (48,3%) do feminino, diferença estatisticamente significativa ($\chi^2(1) = 5,574$; $p = 0,018$), sendo que as alunas apresentaram maior dificuldade na execução do teste 2 de ritmo. No teste 3, o número de adolescentes que não conseguiram executar o teste no ritmo foi muito alto, 118 (79,2%), sendo 45 (72,6%) do sexo masculino e 73 (83,9%) do sexo feminino, diferença marginalmente significativa ($\chi^2(1) = 2,819$; $p = 0,093$), de forma que os testes de ritmo dependem significativamente do sexo. Por outro lado, não se constatou diferença significativa, ($p > 0,25$), em qualquer dos três testes de ritmo com relação às variáveis peso e altura, de maneira que andar no ritmo não depende do peso ou da altura do adolescente.

4.4 COMPARAR MAIS DE DUAS MÉDIAS COM PRESSUPOSTO DE NORMALIDADE

Hipótese: Os alunos de Dourados, com mais idade, executaram o teste das três faixas mais rapidamente do que os de menos idade.

Experimento: Anotar o tempo para executar o teste das três faixas; **Teste:** análise de variância (ANOVA), e teste de Tukey para comparação múltipla de médias, (BUSSAB, 1986).

Procedimento: Abra no SPSS a Planilha Lateralidade. {Análise> Comparar médias> Teste ANOVA para amostras independentes> Indicar como variável de teste, o tempo de execução do teste das três faixas e como variável categórica, a idade> Opções, marcar descritiva e

teste de homogeneidade de variância, clicar continue> Teste a posteriori de comparações múltiplas, marcar Tukey. Clicar em continue e em OK}.

Justificativa: As variáveis envolvidas são tempo de execução (dependente, quantitativa) e idade do aluno (independente, categórica, com 3 grupos: 9, 10 e 11 anos). Essa variável também tem grupos excludentes, pois o aluno que está com nove anos não pode estar no grupo dos que tem 10 ou 11 anos. Para este estudo, como já foi dito, a variável tempo de execução apresentou-se normal, permitindo a utilização da ANOVA. Caso isso não ocorresse, seria necessário aplicar a prova de Kruskal-Wallis (prova não paramétrica). Também, foi verificada a igualdade de variâncias do tempo de execução do teste das três faixas entre os três anos de escolaridade (Teste de Levene). Caso isso não ocorresse, seria necessário observar o tamanho da amostra de cada grupo: se fosse grande, poder-se-ia continuar, do contrário, ter-se-iam que procurar por transformações adequadas da variável tempo ou por testes não paramétricos. Uma forma simples de verificar e entender o conceito de igualdade de variâncias é fazer o gráfico de box-plot. Caixas de tamanhos diferentes, com valores extremos e outliers podem estar indicando que as variâncias dos grupos são diferentes.

Como pode ser observado na Tabela 4.4.1, não houve diferença significativa no tempo necessário para execução do teste das três faixas, em relação aos três grupos de idade, o que não exige a aplicação de um teste complementar, para verificar quais grupos são diferentes (teste de Tukey).

Tabela 4.4.1 - Distribuição das médias do tempo necessário para execução do Teste das três faixas de acordo com a idade do aluno.

Idade [anos]	Número adolescentes	Média [sg]	Desvio padrão	Estatística	P-valor
Nove	79	19,77	4,41	$F_{(2,141)} = 2,30$	0,104
Dez	59	18,22	3,83		
Onze	6	18,82	5,20		

Apresentação do resultado da prova Teste das três faixas, no texto:

$F(gln, gld)$ = valor de F; p = valor da significância da amostra.

Onde,

gln = graus de liberdade do numerador (igual ao número de tratamentos ou grupos menos 1)

gld = graus de liberdade do denominador (igual ao graus de liberdade do resíduo)

Conclui-se que, em média, o tempo necessário para a execução do teste das três faixas não apresenta diferença significativa, ($F_{(2,141)} = 2,30$; $p = 0,104$), entre os alunos de 9, 10 e 11 anos em Dourados, MS.

4.5 EXISTÊNCIA DE CORRELAÇÃO ENTRE DUAS VARIÁVEIS

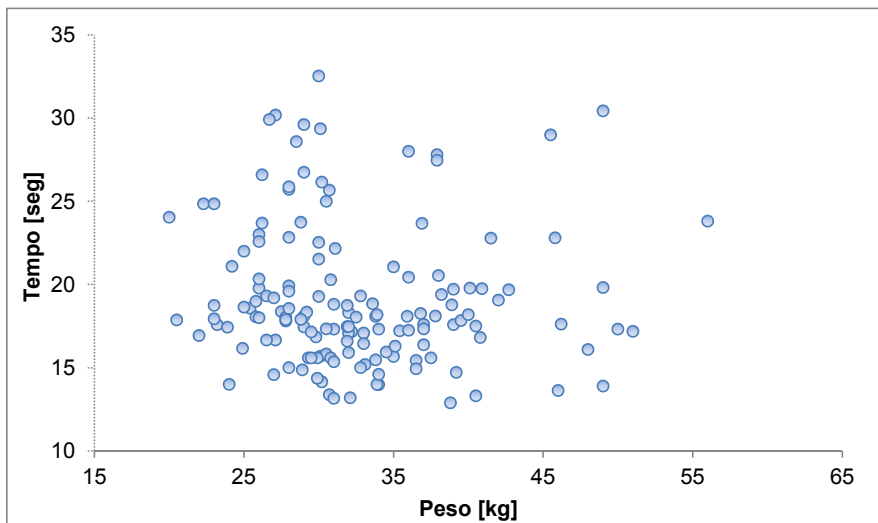
Hipótese: Existe uma relação do peso dos alunos com o tempo de execução do teste das três faixas. A escolha dos testes permite ressaltar a natureza dos dados estatísticos a serem apresentados e sua relação com as sucessivas afirmações contidas nas hipóteses.

Experimento: Anotar o tempo que o aluno precisou para executar o teste das três faixas e seu respectivo peso; **Teste:** Correlação de Pearson, (DAWSON; TRAPP, 2003).

Procedimento: Abra no SPSS a Planilha Lateralidade. {Entre no quadro de Análise> Correlação> Bivariada> Selecionar as variáveis de teste, o tempo de execução do teste das três faixas e o peso do aluno. Clicar OK}.

Justificativa: As variáveis envolvidas são peso (quantitativa) e tempo de execução (quantitativa). Ambas as variáveis apresentaram distribuição normal. Para a correlação, caso as variáveis não tivessem sido normais ou a amostra fosse pequena, seria necessário aplicar a correlação de Spearman.

Gráfico 4.5.1 - Distribuição do tempo para execução do teste em função do peso do estudante.



Fonte: elaborado a partir de dados do apêndice.

O **coeficiente de correlação de Pearson** verifica se existe relação linear entre duas variáveis quantitativas. Quando essa associação é alta, procura-se um modelo que a explique, que é a regressão. Neste estudo, o coeficiente de correlação foi de $r_{(149)} = -0,071$, com um $p = 0,388$, ou seja, o valor da correlação foi baixo e não significativo. Sugestão: sempre faça um gráfico (diagrama de pontos ou scatter plot), a fim de inspecionar visualmente se a relação é linear, pois, caso exista uma relação não linear (como por exemplo, exponencial, quadrática etc.), o coeficiente de correlação de Pearson será baixo e você poderá concluir erradamente que não existe relação, quando, na realidade, não existe relação linear.

Apresentação do resultado da correlação no texto:

$$r_{(\text{número de pares})} = \text{valor da correlação}; p \text{ valor.}$$

Em nossa pesquisa: $r_{(148)} = -0,071$, $p = 0,388$. Portanto o valor da correlação é baixo e não difere significativamente de zero, descartando-se a hipótese de que a obesidade possa ser um fator interferente no tempo de execução do teste.

A **regressão linear** é uma técnica utilizada para se ajustar uma reta a um conjunto de dados, de tal forma que esse modelo possa ser utilizado para explicar essa relação, bem como para prever dados, tanto dentro do intervalo amostrado, quanto fora dele. Caso não seja possível ajustar uma reta, ou seja, os dados não estejam alinhados linearmente, é necessário linearizar a relação, usando-se transformações (funções, tais como a exponencial, a logarítmica, entre outras), ou os modelos de regressão não linear. Observação: no exemplo acima, parece não se aplicar qualquer ajuste aos dados, porque eles não apresentam alguma tendência.

Na Figura 4.5.1, pode-se observar que os dados (pontos) se aproximam de uma reta, que foi determinada pela análise de regressão linear. Considerando X = Peso dos alunos (variável independente) e Y = Tempo para execução do teste das três faixas (variável dependente), podemos formular a seguinte hipótese:

4.6 EXISTÊNCIA DE UMA RELAÇÃO DE DEPENDÊNCIA ENTRE VARIÁVEIS

Hipótese: O tempo de execução do teste das três faixas depende linearmente do peso dos alunos.

Experimento: Anotar o tempo que o aluno precisou para executar o teste das três faixas e seu respectivo peso; **Teste:** análise de regressão para verificar se existe uma relação de dependência linear, (DAWSON; TRAPP, 2003).

Procedimento: Abra no SPSS a Planilha Lateralidade. {Análise> Regressão> Linear, (Indicar como variável dependente o tempo de execução do teste das três faixas e como independente a variável o peso dos alunos), clicar OK}.

Neste caso, a equação da reta foi $Y=20,62-0,045X$, este é o formato para apresentação no texto, ou seja, para cada 1 kg a mais na escala do peso, o aluno diminui o seu tempo de execução do teste em 0,045 segundo (ver Figura 4.6.1). Observe na Tabela 4.6.1 que

o coeficiente angular da reta estimada não difere significativamente de zero, ($t = -0,866$; $p = 0,388$), portanto, a melhor estimativa para a variável dependente é $Y = 20,62$, isto é independe de X .

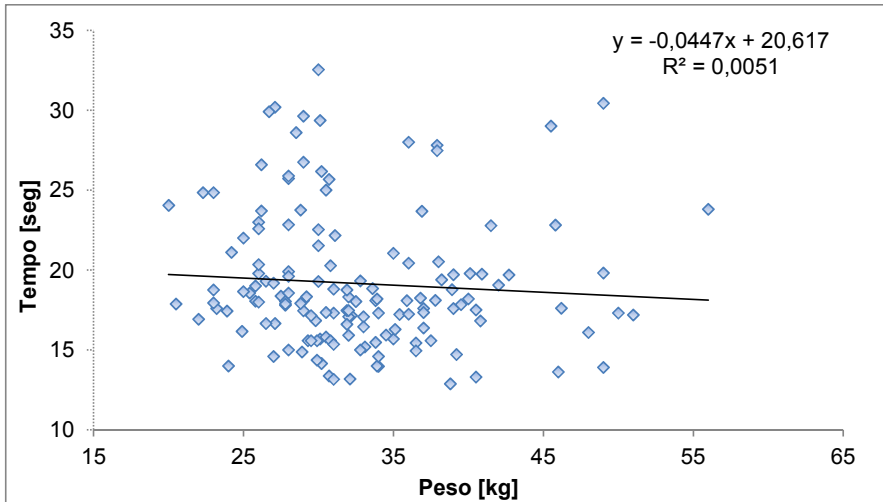
No caso de o coeficiente angular ser significativamente diferente de zero, seria necessário apresentar também o valor do **coeficiente de determinação** (R^2), que determina o quanto da variação da variável dependente é explicada pela variação da variável independente. Neste caso, o coeficiente de determinação foi de 0,51%, ou seja, muito baixo. Isso quer dizer que 0,51% do tempo necessário para executar o teste pode ser explicado pelo peso e 99,5% do desempenho, por outras variáveis. Para apresentação no texto esse coeficiente é, $R^2 = 0,005\%$.

Tabela 4. 6.1 - Relatório de saída do aplicativo.

Modelo	Coeficientes no estandarizados(a)		Coeficientes tipificados		Sig.
	B	Error típ.	Beta	t	
1 (Constante)	20,621	1,723			
Peso [kg]	-0,045	0,052	-0,071	-0,866	0,388

a. Variável dependente: Tempo [seg]

Gráfico 4.6.1 - Relação entre peso do estudante e tempo gasto para executar o teste das três faixas.



Fonte: elaborado a partir de dados do apêndice.

O ideal é apresentar o gráfico de dispersão com a reta ajustada, que possibilita ao leitor observar como os dados estão em relação à reta, bem como escrever em um parágrafo a equação da reta e o coeficiente de determinação, interpretando esses parâmetros.

Quanto à apresentação da figura no texto, esta deve:

- 1) Ter o nome dos eixos e a unidade de medida utilizada;
- 2) Ter um título logo acima da figura e o título deve explicar a figura.

Como sugestão, o ideal é inseri-la em uma caixa de texto, para não ter problemas com o arquivo do Word. Formatar caixa de texto, sem linha e sem preenchimento. Outra sugestão é que, caso você queira editar seu gráfico para alguma modificação, é preferível deletar o gráfico do arquivo Word, editar o gráfico no SPSS ou Excel (para isso, salve todos os seus gráficos no Excel ou SPSS), fazer as correções e colar de novo. Por essa razão, coloque sempre o título via Word, e não via Excel ou SPSS ou qualquer outro editor gráfico.

CONCLUSÃO

O desenvolvimento da disciplina “Bioestatística e Ensaio do Movimento Humano” possibilitou o exercício aplicativo de conceitos e testes sobre dados coletados durante o curso, que farão parte da vida profissional dos estudantes de Educação Física Licenciatura da Plataforma Freire, da UEMS, bem como de outras Instituições de Ensino.

As relações e correlações que permitiram a obtenção dos resultados constituem o cerne da análise e da interpretação de resultados, e oferecem evidências relevantes na verificação das hipóteses, confirmando ou negando-as, constituindo, assim, a parte mais importante de um relatório.

Ao evidenciar as conquistas, cabe apontar as relações entre os fatos e a teoria e os modelos que se unem e se completam.

REFERÊNCIAS

- BOLFARINE, H.; BUSSAB, W. O. **Elementos de Amostragem**. São Paulo: Blucher, 1 ed. 2005.
- BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva, 8 ed. 2011.
- BUSSAB, W. O. **Análise de variância e de regressão: uma introdução**. São Paulo: Atual, 1986.
- DAWSON, B.; TRAPP, R. G. **Bioestatística básica e clínica**. Rio de Janeiro: McGraw-Hill Internacional do Brasil, 3 ed. 2003.
- FONSECA, J. S.; MARTINS, G. A. **Curso de estatística**. São Paulo: Atlas, 5 ed., 1994.
- FRAISSE P.; PIAGET, J. **Tratado de psicologia experimental**. Rio de Janeiro: Forense, 2 ed. Vol. I, 1972.
- GOLDSBERGER, M. **Research on teaching in physical education: a commentary on Silverman's review**. *Research quarterly exercise sport*. 62(4): 352-364. 1991.
- LEVIN, J. **Estatística aplicada às ciências humanas**. São Paulo: Harper & Row do Brasil, 1985.
- LIKERT, Rensis. A Technique for the Measurement of Attitudes. **Archives of Psychology**. 1932.
- MARGOTTO, P. R. **Estatística computacional uso do SPSS**, Brasília, 2014. Disponível em www.paulomargotto.com.br, acessado em 29/07/2014.
- MEYER, P. **Probabilidade: aplicação e estatística**. Rio de Janeiro: Livros Técnicos e Científicos, 1982.
- MILLER, D. C. **Handbook of research design and social measurement**. Newbury Park, Calif.: Sage Publications, 5 ed. 1991.
- MOREIRA, Wagner Wey. **Educação física & esporte: perspectivas para o século XXI**. São Paulo: Papyrus, 1993.

SIEGEL, S. **Estatística não paramétrica para as ciências do comportamento**. São Paulo: McGraw Ltda, 1981.

SILVA, N. N. **Amostragem probabilística: um curso introdutório**. São Paulo: edusp, 1998.

SPSS versão 21 (SPSS IBM, New York, EUA) foi utilizado para realizar a análise estatística.

STEFANELO, Lourdes Lago; FACCENDA, Odival. **Lateralidade e Ritmo por Gêneros em Estudantes Pré-Adolescentes**. The FIEP Bulletin, v.81, p.385 - 390, 2011.

WONNACOTT, R. J.; WONNACOTT, T. H. **Introdução a Estatística**. Rio de Janeiro: Livros Técnicos e Científicos, 1980.

Anexo A - Tábuas Estatísticas

Tábua I - Valores críticos - Normal

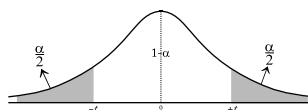
Tábua II - Valores críticos - t de student

Tábua III - Valores críticos - Qui-quadrado

Tábua IV - Valores críticos - F de Snedecor

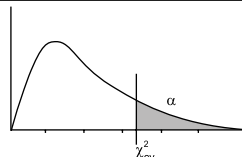
Tábua V - Números Aleatórios

Tábua II - Valores críticos - t de student
 α : Nível de significância;
 ρ : Graus de liberdade.



$\rho \backslash \alpha$	0,500	0,200	0,100	0,050	0,020	0,010	0,002
1	1,000	3,078	6,314	12,706	31,821	63,657	318,31
2	0,816	1,886	2,920	4,303	6,965	9,925	22,327
3	0,765	1,638	2,353	3,182	4,451	5,841	10,214
4	0,741	1,533	2,132	2,776	3,747	4,604	7,173
5	0,727	1,476	2,015	2,571	3,365	4,032	5,893
6	0,718	1,440	1,943	2,447	3,143	3,707	5,208
7	0,711	1,415	1,895	2,365	2,998	3,499	4,785
8	0,706	1,397	1,860	2,306	2,896	3,355	4,501
9	0,703	1,383	1,833	2,262	2,821	3,250	4,297
10	0,700	1,372	1,812	2,228	2,764	3,169	4,144
11	0,697	1,363	1,796	2,201	2,718	3,106	3,025
12	0,695	1,356	1,782	2,179	2,681	3,055	3,930
13	0,694	1,350	1,771	2,160	2,650	3,012	3,852
14	0,692	1,345	1,761	2,145	2,624	2,977	3,787
15	0,691	1,341	1,753	2,131	2,602	2,947	3,733
16	0,690	1,337	1,746	2,120	2,583	2,921	3,686
17	0,689	1,333	1,740	2,110	2,567	2,898	3,646
18	0,688	1,330	1,734	2,101	2,552	2,878	3,610
19	0,688	1,328	1,729	2,093	2,539	2,861	3,579
20	0,687	1,325	1,725	2,086	2,528	2,845	3,552
21	0,686	1,323	1,721	2,080	2,518	2,831	3,527
22	0,686	1,321	1,717	2,074	2,508	2,819	3,505
23	0,685	1,319	1,714	2,069	2,500	2,807	3,485
24	0,685	1,318	1,711	2,064	2,492	2,797	3,467
25	0,684	1,316	1,708	2,060	2,485	2,787	3,450
26	0,684	1,315	1,706	2,056	2,479	2,779	3,435
27	0,684	1,314	1,703	2,052	2,473	2,771	3,421
28	0,683	1,313	1,701	2,048	2,467	2,763	3,408
29	0,683	1,311	1,699	2,045	2,462	2,756	3,396
30	0,683	1,310	1,697	2,042	2,457	2,750	3,385
60	0,679	1,296	1,671	2,000	2,390	2,660	3,232
120	0,677	1,289	1,658	1,980	2,358	2,617	3,160
∞	0,674	1,282	1,645	1,960	2,326	2,576	3,090

Tábua III - Valores críticos - Qui-quadrado.
Valores de X tais que $p(X \geq x) = \alpha$



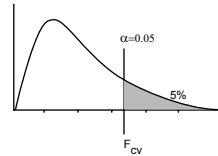
n/α	0,995	0,990	0,975	0,950	0,900	0,750	0,500	0,250	0,100	0,050	0,025	0,010	0,005
1	0,0000	0,0002	0,0010	0,0039	0,0158	0,1015	0,4549	1,3233	2,7055	3,8415	5,0239	6,6349	7,8794
2	0,0100	0,0201	0,0506	0,1026	0,2107	0,5754	1,3863	2,7726	4,6052	5,9915	7,3778	9,2103	10,5966
3	0,0717	0,1148	0,2158	0,3118	0,5844	1,2125	2,3660	4,1083	6,2514	7,8147	9,3484	11,3449	12,8382
4	0,2070	0,2971	0,4844	0,7107	1,0636	1,9226	3,3567	5,3853	7,7794	9,4877	11,1433	13,2767	14,8603
5	0,4117	0,5543	0,8312	1,1455	1,6103	2,6746	4,3515	6,6257	9,2364	11,0705	12,8325	15,0863	16,7496
6	0,6557	0,8721	1,2373	1,6354	2,2041	3,4546	5,3481	7,8408	10,6446	12,5916	14,4494	16,8119	18,5476
7	0,9893	1,2390	1,6889	2,1673	2,8331	4,2549	6,3458	9,0371	12,0170	14,0671	16,0128	18,4753	20,2777
8	1,3444	1,6465	2,1797	2,7326	3,4895	5,0706	7,3441	10,2189	13,3616	15,5073	17,5345	20,0902	21,9550
9	1,7349	2,0879	2,7004	3,3251	4,1682	5,8988	8,3428	11,3888	14,6837	16,9190	19,0228	21,6660	23,5894
10	2,1559	2,5582	3,2470	3,9403	4,8652	6,7372	9,3418	12,5489	15,9872	18,3070	20,4832	23,2093	25,1882
11	2,6032	3,0535	3,8157	4,5748	5,5778	7,5841	10,3410	13,7007	17,2750	19,6751	21,9200	24,7250	26,7568
12	3,0738	3,5706	4,4038	5,2260	6,3038	8,4384	11,3403	14,8454	18,5493	21,0261	23,3367	26,2170	28,2995
13	3,5650	4,1069	5,0088	5,8919	7,0415	9,2991	12,3398	15,9839	19,8119	22,3620	24,7356	27,6882	29,8195
14	4,0747	4,6604	5,6287	6,5706	7,7895	10,1653	13,3393	17,1169	21,0641	23,6848	26,1189	29,1412	31,3193
15	4,6009	5,2293	6,2621	7,2609	8,5468	11,0365	14,3389	18,2451	22,3071	24,9958	27,4884	30,5779	32,8013
16	5,1422	5,8122	6,9077	7,9616	9,3122	11,9122	15,3385	19,3689	23,5418	26,2962	28,8454	31,9999	34,2672
17	5,6972	6,4078	7,5642	8,6718	10,0852	12,7919	16,3382	20,4887	24,7690	27,5871	30,1910	33,4087	35,7185
18	6,2648	7,0149	8,2307	9,3905	10,8649	13,6753	17,3379	21,6049	25,9894	28,8693	31,5264	34,8053	37,1565
19	6,8440	7,6327	8,9065	10,1170	11,6509	14,5620	18,3377	22,7178	27,2036	30,1435	32,8523	36,1909	38,5823
20	7,4338	8,2604	9,5908	10,8508	12,4426	15,4518	19,3374	23,8277	28,4120	31,34104	34,1696	37,5662	39,9968
21	8,0337	8,8972	10,2829	11,5913	13,2696	16,3444	20,3372	24,9348	29,6151	32,6706	35,4789	38,9322	41,4011
22	8,6427	9,5425	10,9823	12,3380	14,0415	17,2396	21,3370	26,0393	30,8133	33,9244	36,7807	40,2894	42,7957
23	9,2604	10,1957	11,6886	13,0905	14,8480	18,1373	22,3369	27,1413	32,0069	35,1725	38,0756	41,6384	44,1813
24	9,8862	10,8564	12,4012	13,8484	15,6587	19,0373	23,3367	28,2412	33,1962	36,4150	39,3641	42,9798	45,5585
25	10,5197	11,5240	13,1197	14,6114	16,4734	19,9393	24,3366	29,3389	34,3816	37,6525	40,6465	44,3141	46,9279
26	11,1602	12,1981	13,8439	15,3792	17,2919	20,8434	25,3365	30,4346	35,5632	38,8851	41,9232	45,6417	48,2899
27	11,8076	12,8785	14,5734	16,1514	18,1139	21,7494	26,3363	31,5284	36,7412	40,1133	43,1945	46,9629	49,6449
28	12,4613	13,5647	15,3079	16,9279	18,9392	22,6572	27,3362	32,6205	37,9159	41,3371	44,4608	48,2782	50,9934
29	13,1211	14,2565	16,0471	17,7084	19,7677	23,5666	28,3361	33,7109	39,0875	42,5570	45,7223	49,5879	52,3356
30	13,7867	14,9535	16,7908	18,4927	20,5992	24,4776	29,3360	34,7997	40,2560	43,7730	46,9792	50,8922	53,6720

Tábua IV - Valores críticos - F de Snedecor.

$P(X \geq x) = 0,05.$

ρ_1 : Graus de liberdade numerador.

ρ_2 : Graus de liberdade denominador.



$\rho_2 \backslash \rho_1$	1	2	3	4	5	6	7	8	9	10	20	30	40	50	100	120	∞
1	161,5	200	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	248,0	250	251,1	251,8	253,0	253,3	254,3
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,45	19,46	19,47	19,48	19,49	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,62	8,59	8,58	8,55	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,75	5,72	5,70	5,66	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,50	4,46	4,44	4,41	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,81	3,77	3,75	3,71	3,7	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,38	3,34	3,32	3,27	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	3,08	3,04	3,02	2,97	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,86	2,83	2,80	2,76	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,77	2,70	2,66	2,64	2,59	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,65	2,57	2,53	2,51	2,46	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,54	2,47	2,43	2,40	2,35	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,46	2,38	2,34	2,31	2,26	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,39	2,31	2,27	2,24	2,19	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,33	2,25	2,20	2,18	2,12	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,28	2,19	2,15	2,12	2,07	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,23	2,15	2,10	2,08	2,02	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,19	2,11	2,06	2,04	1,98	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,16	2,07	2,03	2,00	1,94	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	2,04	1,99	1,97	1,91	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,10	2,01	1,96	1,94	1,88	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,07	1,98	1,94	1,91	1,85	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,05	1,96	1,91	1,88	1,82	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,03	1,94	1,89	1,86	1,80	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,01	1,92	1,87	1,84	1,78	1,77	1,71
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	1,93	1,84	1,79	1,76	1,70	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	1,84	1,74	1,69	1,66	1,59	1,58	1,51
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,03	1,78	1,69	1,63	1,60	1,52	1,51	1,44
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,75	1,65	1,59	1,56	1,48	1,47	1,39
70	3,98	3,13	2,74	2,50	2,35	2,23	2,14	2,07	2,02	1,97	1,72	1,62	1,57	1,53	1,45	1,44	1,35
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,70	1,60	1,54	1,51	1,43	1,41	1,32
90	3,95	3,10	2,71	2,47	2,32	2,20	2,11	2,04	1,99	1,94	1,69	1,59	1,53	1,49	1,41	1,39	1,30
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,68	1,57	1,52	1,48	1,39	1,38	1,28
120	3,92	3,07	2,68	2,45	2,29	2,18	2,09	2,02	1,96	1,91	1,66	1,55	1,50	1,46	1,37	1,35	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,57	1,46	1,39	1,35	1,24	1,22	1,00

Tábua V - Números Aleatórios

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

Anexo B - Escala de Sono de Epworth Versão em Português no Brasil (ESS/BR)

Qual a probabilidade de você cochilar ou dormir, e não apenas se sentir cansado, nas seguintes situações? Considere o modo de vida que você tem levado recentemente. Mesmo que você não tenha feito algumas destas coisas recentemente, tente imaginar como elas o afetariam. Escolha o número mais apropriado para responder cada questão:

0= nunca cochilaria

1= pequena probabilidade de cochilar

2= média probabilidade de cochilar

3= grande probabilidade de cochilar

Situação	Probabilidade de Cochilar			
S1: Sentado e lendo	0	1	2	3
S2: Assistindo à TV	0	1	2	3
S3: Sentado, quieto, em um lugar público (por exemplo, em um teatro, reunião ou palestra)	0	1	2	3
S4: Andando de carro por uma hora sem parar, como passageiro	0	1	2	3
S5: Ao deitar-se à tarde para descansar, quando possível	0	1	2	3
S6: Sentado conversando com alguém	0	1	2	3
S7: Sentado quieto após o almoço sem bebida de álcool	0	1	2	3
S8: Em um carro parado no trânsito por alguns minutos	0	1	2	3

Como Avaliar as Questões do Sono

Some todos os pontos de cada respondente e conclua. Se o Escore total for:

Até 6 Pontos: Parabéns, você tem um sono saudável.

de 7 a 9 Pontos: Luz amarela, seu sono pode não ser adequado.

Acima de 10 Pontos: Alerta Vermelho! Você pode estar em risco.

Apêndice A - Planilha Ritmo e Lateralidade

Id: identificador do indivíduo coletado; M: Município da coleta (1 = Dourados, 2 = Jardim); ESC: Escolas (Joaquim Murtinho, Aurora Pedroso, Clarice Bastos, Franklin Azambuja e Etalívio Penzo); Ida: Idade do aluno; Sx: Sexo do aluno (0 = Feminino; 1 = Masculino); AltM: altura média referente a três medidas; DT: Número de Desvios no teste de ritmo em três velocidades, T1 = Tempo 64 (vel.12), T2 = Tempo 120 (vel.22), T3 = Tempo 184 (vel. 32); T: Tempo para execução do teste das três faixas; ED: Como iniciou o teste (1 = Esquerda/Direita, 2 = Direita/Esquerda); REIN: Quantas vezes reiniciou o teste.

Id	M	Esc	Ida	Sx	Peso	AltM	DT1	DT2	DT3	T	ED	REIN
1	1	1	10	0	40,8	1,440	3	6	3	16,82	1	0
2	1	1	11	0	30,1	1,425	6	6	6	15,69	1	0
3	1	1	10	0	45,8	1,443	6	6	6	22,81	1	0
4	1	1	11	0	35,4	1,462	6	6	6	17,22	1	0
5	1	1	10	0	41,5	1,485	1	6	6	22,78	1	0
6	1	1	10	0	35,9	1,410	6	6	6	18,09	2	1
7	1	1	9	1	27,1	1,310	0	1	0	16,66	1	1
8	1	1	11	1	39,2	1,448	3	3	6	14,72	1	0
9	1	1	10	1	35,0	1,443	0	0	3	15,68	1	0
10	1	1	10	1	25,4	1,329	6	0	6	18,57	1	1
11	1	2		0	20,0	1,357	1	3	6	24,05	1	2
12	1	2	9	0	27,0	1,350	1	6	6	14,59	1	0
13	1	2	9	0	26,0	1,346	0	0	6	19,78	2	0
14	1	2	9	0	29,3	1,395	1	2	6	15,59	2	0
15	1	2	9	0	42,7	1,392	6	6	6	19,69	2	0
16	1	2	9	0	30,7	1,446	0	6	6	25,68	2	0
17	1	2	10	0	38,2	1,511	3	6	6	19,4	2	0
18	1	2	10	0	30,7	1,346	1	0	1	13,38	1	0
19	1	2	9	0	25,8	1,323	0	0	6	19	2	0
20	1	2	10	0	40,5	1,445	1	2	6	17,5	2	0
21	1	2	10	0	34,0	1,423	6	6	1	13,97	2	0
22	1	2		0	26,2	1,361	6	6	6	23,69	2	0
23	1	2	9	0	30,0	1,438	6	0	6	21,53	1	1
24	1	2	10	0	37,0	1,473	6	0	1	16,37	2	0
25	1	2	9	0	50,0	1,407	6	0	6	17,31	1	0
26	1	2	9	0	25,0	1,365	6	6	6	22	1	3

Apêndice A - continuação

Id	M	Esc	Ida	Sx	Peso	AltM	DT1	DT2	DT3	T	ED	REIN
27	1	2	10	0	31,1	1,450	6	2	6	22,16	1	0
28	1	2	9	0	28,0	1,315	6	0	6	22,84	1	0
29	1	2	10	1	24,2	1,261	1	0	6	21,1	2	0
30	1	2	9	1	33,6	1,356	0	0	1	18,85	1	0
31	1	2	9	1	29,0	1,291	6	2	3	17,45	1	0
32	1	2	9	1	32,2	1,393	6	6	6	17,16	1	0
33	1	2	9	1	27,5	1,345	0	0	2	18,37	1	1
34	1	2	9	1	32,0	1,290	6	1	6	17,09	1	2
35	1	2		1	32,0	1,325	0	0	1	18,31	2	1
36	1	2	10	1	24,0	1,285	0	6	6	14	2	0
37	1	2	9	1	26,2	1,332	0	2	6	26,6	2	0
38	1	2	10	1	30,0	1,303	6	0	6	19,28	1	0
39	1	3	10	0	32,5	1,460	6	2	6	18,03	2	0
40	1	3	9	0	27,1	1,394	6	0	5	30,19	1	1
41	1	3	10	0	40,1	1,393	6	0	6	19,78	1	2
42	1	3	10	0	31,9	1,390	6	2	6	17,47	1	0
43	1	3	9	0	46,2	1,483	2	0	6	17,62	2	0
44	1	3	10	0	30,2	1,450	5	5	6	14,15	2	0
45	1	3	9	0	23,2	1,302	6	0	6	17,59	2	0
46	1	3	9	0	26,5	1,361	6	6	6	19,31	1	0
47	1	3	10	0	33,1	1,430	6	1	6	15,19	1	0
48	1	3	9	0	37,0	1,529	0	0	6	17,6	2	0
49	1	3	9	0	29,0	1,343	6	6	2	18,09	2	0
50	1	3	9	0	28,9	1,333	6	0	6	14,87	1	0
51	1	3	9	0	33,8	1,480	1	0	6	18,09	1	0
52	1	3	9	0	28,8	1,307	2	6	6	23,75	1	1
53	1	3	10	0	32,1	1,410	6	6	2	13,19	2	1
54	1	3	9	0	33,0	1,346	6	0	6	17,09	2	0
55	1	3	9	0	30,2	1,436	6	6	6	26,16	2	2
56	1	3	10	0	23,9	1,230	6	6	2	17,44	1	0
57	1	3	9	0	22,0	1,217	6	2	6	16,93	1	1
58	1	3	10	1	56,0	1,548	6	0	6	23,81	1	2
59	1	3	10	1	36,5	1,497	6	0	6	15,44	1	1
60	1	3	10	1	27,8	1,341	6	1	6	18,03	1	1
61	1	3	10	1	29,2	1,341	6	1	6	18,34	1	0
62	1	3	10	1	33,9	1,415	6	1	6	14	2	1
63	1	3	9	1	30,0	1,394	6	1	6	22,54	1	0
64	1	3	10	1	31,9	1,403	6	0	6	18,75	2	0
65	1	3	10	1	37,8	1,334	6	0	1	18,1	1	3
66	1	3	9	1	30,5	1,422	0	6	6	15,82	2	3

Apêndice A - continuação

Id	M	Esc	Ida	Sx	Peso	AltM	DT1	DT2	DT3	T	ED	REIN
67	1	3	9	1	28,0	1,346	6	0	6		1	0
68	1	4	10	0	29,9	1,371	6	1	6	15,59	1	1
69	1	4	10	0	30,1	1,344	6	1	6	29,37	2	0
70	1	4	10	0	37,9	1,505	6	0	6	27,81	1	1
71	1	4	10	0	33,9	1,397	6	2	2	18,19	1	0
72	1	4	9	0	27,8	1,353	6	6	6	17,81	2	2
73	1	4	9	0	28,8	1,304	6	6	1	17,91	1	2
74	1	4	9	0	31,0	1,316	6	2	6	18,81	2	2
75	1	4	10	0	39,0	1,365	6	3	6	17,59	2	2
76	1	4	9	0	30,8	1,427	1	6	6	20,28	2	2
77	1	4	9	0	29,0	1,367	6	1	6	29,62	1	2
78	1	4	9	0	29,8	1,375	6	6	0	16,84	1	3
79	1	4	10	0	36,8	1,408	6	0	6	18,25	2	0
80	1	4	9	0	33,8	1,416	5	0	6	15,47	1	2
81	1	4	10	0	30,8	1,376	3	1	6	15,59	1	0
82	1	4	9	0	36,9	1,452	6	6	6	23,68	1	1
83	1	4	10	0	29,5	1,382	6	3	6	15,59	2	0
84	1	4	9	0	25,0	1,349	6	6	6	18,65	2	0
85	1	4	9	0	34,5	1,478	6	2	6	15,94	1	0
86	1	4	10	0	49,0	1,340	6	6	6	30,44	1	2
87	1	4	9	0	26,7	1,323	6	0	6	29,91	1	1
88	1	4	10	0	38,9	1,449	6	6	6	18,78	2	1
89	1	4	9	1	28,0	1,352	0	0	6	25,72	2	2
90	1	4	10	1	31,0	1,365	5	6	3	17,31	2	1
91	1	4	9	1	29,9	1,347	6	1	6	14,37	1	1
92	1	4	10	1	51,0	1,471	6	6	6	17,19	1	1
93	1	4	9	1	31,9	1,452	6	6	6	16,6	2	1
94	1	4	10	1	35,1	1,421	6	1	6	16,29	1	2
95	1	4	9	1	34,0	1,326	6	2	6	14,6	2	2
96	1	4	10	1	31,0	1,403	6	3	2	15,35	1	1
97	1	4		1	40,9	1,377	2	6	6	19,75	1	3
98	1	4	9	1	26,5	1,332	3	0	6	16,66	1	1
99	1	4	9	1	38,0	1,415	6	1	6	20,53	1	3
100	1	4	9	1	25,8	1,305	6	2	6	18,06	2	3
101	1	4	9	1	22,3	1,327	6	2	6	24,85	2	1
102	1	4	10	1	27,8	1,375	0	6	6	17,88	1	3
103	1	4	9	1	37,5	1,413	6	1	1	15,59	2	2
104	1	5	9	1	28,0	1,373	0	6	6	19,91	1	1
105	1	5	9	0	36,0	1,340	0	6	6	28	2	0
106	1	5	10	1	30,5	1,340	6	1	1	25	2	0

Apêndice A - continuação

Id	M	Esc	Ida	Sx	Peso	AltM	DT1	DT2	DT3	T	ED	REIN
107	1	5	9	0	26,0	1,340	1	0	2	23	1	0
108	1	5	10	1	28,0	1,330	0	6	6	25,88	1	0
109	1	5	9	1	39,5	1,317	1	0	1	17,84	2	0
110	1	5	10	1	32,0	1,270	6	6	6	17,47	2	0
111	1	5	10	0	36,0	1,290	6	6	6	20,44	1	2
112	1	5	10	1	24,9	1,390	1	6	6	16,16	1	1
113	1	5	9	0	36,5	1,360	1	1	3	14,94	2	0
114	1	5	10	0	39,0	1,490	6	1	6	19,72	1	0
115	1	5	9	0	26,0	1,273	6	0	6	22,59	1	2
116	1	5	9	0	37,9	1,280	6	6	1	27,47	1	1
117	1	5	9	0	26,0	1,265	6	6	6	20,34	1	0
118	1	5	9	1	23,0	1,270	6	0	6	24,85	1	0
119	1	5	9	0	30,0	1,395	6	0	6	32,53	1	0
120	1	5	9	1	32,8	1,380	1	0	0	15	1	0
121	1	5	9	0	28,0	1,350	1	6	6	18,56	1	
122	1	5	9	0	26,0	1,340	6	0	6	18	2	
123	1	5	9	0	23,0	1,215	3	0	6	18,75	1	0
124	1	5	9	0	29,0	1,430	6	1	6	26,75	1	2
125	1	5	9	0	40,0	1,415	0	6	6	18,19	1	0
126	1	5	9	1	48,0	1,480	6	6	6	16,1	1	
127	1	5	9	1	28,5	1,375	1	2	6	28,6	2	0
128	1	5	10	1	32,0	1,415	6	1	2	15,91	2	
129	1	5	9	1	46,0	1,390	1	0	6	13,63	1	1
130	1	5	9	1	29,5	1,373	6	6	6	17,15	2	0
131	1	5	10	0	20,5	1,290	6	6	6	17,87	1	0
132	1	5	9	0	32,8	1,390	3	0	0	19,32	1	1
133	1	5	9	0	30,5	1,390	6	6	6	17,34	2	0
134	1	5	9	0	35,0	1,350	6	6	6	21,06	2	
135	1	5	10	0	28,0	1,450	1	6	6	19,6	2	0
136	1	5	10	1	33,0	1,410	3	1	6	16,45	1	1
137	1	5	10	1	23,0	1,320	5	1	1	17,94	1	
138	1	5	9	1	27,0	1,334	2	6	0	19,19	1	3
139	1	5	9	1	37,0	1,410	2	0	6	17,34	1	4
140	1	5	10	0	34,0	1,391	2	3	2	17,31	2	0
141	1	5	11	0	42,0	1,510	6	2	6	19,06	2	2
142	1	5	9	0	49,0	1,390	6	2	1	19,81	1	1
143	1	5	11	1	36,0	1,420	6	0	6	17,25	2	0
144	1	5	11	1	45,5	1,401	0	1	1	29	1	3
145	1	5	10	1	31,0	1,395	0	0	0	13,16	1	0
146	1	5	10	0	40,5	1,480	6	2	6	13,3	1	0

Apêndice A - continuação

Id	M	Esc	Ida	Sx	Peso	AltM	DT1	DT2	DT3	T	ED	REIN
147	1	5	10	1	49,0	1,497	1	0	0	13,9	1	0
148	1	5	9	1	28,0	1,310	0	2	0	15	1	0
149	1	5	9	1	38,8	1,370	6	6	6	12,88	1	1
150	2	6	10	1	37,2	1,445	2	1	4	19	1	3
151	2	6	10	1	32,0	1,440	3	0	2	15,25	1	0
152	2	6	10	0	35,4	1,410	6	6	6	13,16	1	0
153	2	6	10	1	24,4	1,280	5	3	6	16,3	1	1
154	2	6	10	0	31,3	1,270	6	1	3	17,56	1	0
155	2	6	10	0	33,9	1,420	6	2	1	18,78	1	0
156	2	6	10	1	25,4	1,320	6	2	3	16,69	2	0
157	2	6	10	0	25,4	1,320	6	2	3	16,69	1	0
158	2	6	11	1	59,1	1,440	6	2	6	18,35	1	0
159	2	6	10	1	33,7	1,370	4	5	4	18	1	1
160	2	6	10	1	26,6	1,340	3	3	6	16,2	1	1
161	2	6	12	1	20,4	1,280	6	6	6	20,34	2	1
162	2	6	10	1	37,6	1,430	3	1	3	11,31	1	0
163	2	6	10	1	30,5	1,457	3	4	3	18,14	1	0
164	2	6	10	1	32,1	1,450	2	3	3	17,11	1	1
165	2	6	10	0	27,0	1,340	3	3	3	17,54	1	1
166	2	6	10	0	33,1	1,415	2	2	3	28,75	1	0
167	2	6	10	0	39,3	1,410	0	6	3	14,68	1	0
168	2	6	11	1	42,2	1,590	1	0	3	17,57	1	1
169	2	6	11	0	44,1	1,540	1	1	3	14,34	2	0
170												
171	2	6	10	1	45,0	1,450	1	2	5	14,78	1	0
172	2	6	10	1	53,6	1,430	6	3	3	13,97	1	0
173	2	6	10	0	47,6	1,470	5	0	3	15,28	1	0
174	2	6	10	0	34,2	1,370	1	0	3	17,12	1	1
175												
176	2	6	11	1	42,0	1,490	1	0	3	12	1	1
177	2	6	10	1	36,7	1,410	1	1	3	13,28	1	1
178	2	6	10	1	31,2	1,440	1	0	3	14,44	1	0
179	2	6	9	1	49,9	1,400	3	1	3	15,56	1	0
180	2	6	10	1	29,8	1,410	6	6	0	15,88	1	0

Apêndice B - Planilha Escala de Sono

Id: identificador do indivíduo coletado; loc: Local de coleta (1=Bares e restaurantes; 2=Atendimento em geral; 3=Policiamento e vigias e 4=Meio acadêmico); Idad: Idade (anos completos); I_Rc: Idade recodificada (0=menos de 30 anos; 1=30 ou mais anos); Sex: Sexo (0=Masculino; 1=feminino); P1:Escolaridade (1=Fundamental incompleto; 2=médio incompleto; 3=médio completo; 4= Superior incompleto; 5=Superior completo); P2: Horário de início do trabalho (1=00:00 h; 2=01:00 h; 3=02:00 h; 4=04:00 h ou mais); P3: Horário que costuma dormir (1=00:00 h; 2=01:00 h; 3=02:00 h; 4=04:00 h ou mais); P4: Apresenta dificuldade para dormir (1=Sim; 2=Não); P5: Toma algum tipo de remédio para dormir (1=Nunca; 2=Raramente; 3=Às vezes; 4=Repetidamente; 5=Sempre); P6: Apresenta problema com ronco (1=Nunca; 2=Raramente; 3=Às vezes; 4=Sempre); P7: Acorda muito cansado (1=Sim; 2=Não; 3=Outra resposta); P8:Horário que costuma acordar (1=Até 8 h; 2=09 h; 3=10 h; 4=A partir das 11 h); P9: Almoça bem (1=Sim; 2=Não); P10: Casado (1=Sim; 2=Não); S:Escala do sono ver metodologia no final da planilha; Dig: Diagnóstico (1 = Seu sono é saudável; 2 = O seu sono pode não ser adequado; 3 = você pode estar em risco).

Id	Loc	Idad	I_Rc	Sex	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	S1	S2	S3	S4	S5	S6	S7	S8	S_Tot	Dig
1	1	20	0	0	2	2	3	2	1	3	2	2	1	2	2	0	1	3	2	0	2	0	10	3
2	1	21	0	1	3	2	3	1	1	1	1	4	1	2	2	1	3	0	3	0	3	0	12	3
3	1	26	0	0	3	3	4	1	2	5	2	4	1	1	1	0	0	0	0	0	0	0	1	1
4	1	28	0	1	2	1	2	2	1	1	2	1	2	1	2	3	0	0	3	0	0	0	8	2
5	1	23	0	1	3	1	2	2	1	1	2	2	1	2	0	0	0	2	2	0	0	0	4	1
6	1	26	0	1	4	1	3	1	1	1	1	3	2	2	0	2	0	3	1	0	2	0	8	2
7	1	24	0	1	1	1	2	1	3	1	1	2	2	2	3	3	3	3	3	2	2	2	21	3
8	1	22	0	0	2	1	4	2	1	2	3	3	1	2	1	1	1	0	2	0	1	0	6	1
9	1	17	0	1	2	1	2	1	1	1	1	1	1	2	1	0	0	3	2	0	3	0	9	2
10	1	46	1	0	3	1	4	2	1	5	2	2	2	1	2	2	1	1	3	0	3	0	12	3
11	1	24	0	1	3	2	4	1	1	3	1	3	1	2	2	3	0	2	3	0	1	0	11	3
12	1	22	0	0	3	2	3	2	1	1	2	1	1	1	3	0	0	0	3	0	3	0	9	2
13	1	17	0	1	2	2	3	2	1	1	2	1	2	2	0	2	1	2	1	0	2	1	9	2

Apêndice B - continuação

Id	Loc	Idad	I_Rc	Sex	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	S1	S2	S3	S4	S5	S6	S7	S8	S_Tot	Dig
14	1	36	1	0	3	3	4	2	1	5	1	3	2	2	3	0	3	3	2	2	2	0	15	3
15	1	21	0	1	1	2	4	2	3	5	1	4	1	2	1	3	0	3	3	0	0	0	10	3
16	1	37	1	0	3	3	4	2	1	1	1	2	1	1	1	1	1	1	2	1	1	1	9	2
17	2	54	1	0	3	4	4	2	1	2	1	1	2	1	2	0	0	2	0	0	2	0	6	1
18	1	21	0	1	4	3	4	2	1	5	1	4	2	1	3	3	3	0	3	0	3	0	15	3
19	1	30	1	1	3	3	2	2	1	5	2	4	2	1	3	3	3	3	3	0	3	0	18	3
20	2	31	1	0	4	1	2	2	1	3	2	1	2	1	0	0	0	0	3	0	0	0	3	1
21	2	22	0	0	3	4	4	1	1	2	2	4	2	2	1	0	0	0	3	0	3	0	7	2
22	3	29	0	0	4	4	4	2	1	1	2	2	2	2	1	1	0	0	2	0	1	0	5	1
23	3	34	1	0	3	4	4	2	1	3	1	1	2	1	1	1	0	1	2	0	0	0	5	1
24	3	30	1	0	5	4	4	2	1	3	2	1	2	2	0	0	0	0	2	0	3	0	5	1
25	1	32	1	1	1	2	3	2	1	1	1	2	1	2	2	0	0	2	0	0	0	0	4	1
26	1	37	1	0	1	2	3	2	1	3	2	1	2	1	2	2	1	2	3	1	2	0	13	3
27	1	42	1	0	1	1	1	2	1	2	1	1	2	1	0	1	0	0	0	0	1	1	3	1
28	1	23	0	1	3	1	1	2	1	2	1	2	2	2	2	3	0	0	0	0	0	0	5	1
29	2	61	1	1	5	1	1	2	1	2	1	1	2	1	0	3	2	3	2	1	2	0	13	3
30	2	39	1	0	5	1	1	2	2	2	1	4	2	1	2	3	2	1	3	0	3	0	14	3
31	2	24	0	1	3	1	1	2	1	2	1	3	2	2	0	2	0	3	3	0	0	0	8	2
32	2	25	0	1	4	1	1	2	1	2	1	3	2	2	2	2	0	3	3	0	1	0	11	3
33	2	37	1	1	3	1	1	2	1	2	1	2	2	2	0	2	0	0	2	0	2	0	6	1
34	1	38	1	0	3	2	4	2	1	2	1	3	2	2	0	3	0	3	3	0	0	0	9	2
35	1	29	0	0	1	2	4	1	1	1	3	4	1	1	3	3	3	3	3	0	2	0	17	3
36	1	24	0	0	4	3	4	1	1	2	3	4	2	1	1	0	0	0	3	0	0	0	4	1
37	1	57	1	1	1	1	2	2	1	2	1	1	2	2	1	1	0	1	1	0	0	0	4	1
38	1	33	1	0	3	1	2	2	1	2	1	2	2	1	2	1	2	2	3	0	2	0	12	3
39	1	43	1	0	3	3	4	1	5	1	1	2	2	2	0	1	2	2	0	0	0	0	5	1
40	1	23	0	0	2	2	4	2	1	2	3	1	1	2	2	2	0	0	2	0	2	0	8	2
41	1	18	0	0	3	1	4	2	1	2	3	3	2	2	0	2	0	3	3	0	3	0	11	3
42	2	47	1	0	3	1	1	1	1	1	1	3	2	1	3	1	1	3	3	0	2	0	13	3
43	1	19	0	1	3	1	1	2	1	1	2	2	2	2	0	0	2	0	3	0	0	0	5	1
44	1	20	0	1	3	1	2	2	1	1	2	1	2	2	3	3	2	3	3	0	3	0	17	3
45	2	26	0	1	3	1	1	2	1	2	1	4	2	1	1	3	0	2	3	0	1	0	10	3
46	1	36	1	0	3	1	1	1	2	1	1	2	1	1	3	3	0	2	3	0	3	1	15	3
47	1	18	0	1	3	1	1	1	3	1	1	2	2	2	3	1	0	3	3	0	3	0	13	3
48	1	22	0	0	3	1	2	2	1	2	3	4	2	2	0	2	0	2	3	0	2	0	9	2
49	1	17	0	1	1	1	3	1	1	1	3	3	1	2	3	3	0	3	1	0	3	2	15	3
50	2	33	1	0	3	1	1	2	1	1	1	2	2	1	2	2	0	3	3	0	3	0	13	3
51	1	41	1	0	1	4	4	1	1	1	3	4	1	1	0	2	0	0	1	0	1	0	4	1
52	1	48	1	1	5	2	3	1	3	1	1	4	2	1	2	0	2	0	3	0	0	0	7	2

Apêndice B - continuação

Id	Loc	Idad	L_Rc	Sex	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	S1	S2	S3	S4	S5	S6	S7	S8	S_Tot	Dig
53	1	26	0	0	3	1	2	2	1	2	1	2	2	2	1	2	3	1	2	0	1	0	10	3
54	1	17	0	1	2	2	4	1	1	1	3	1	1	2	2	1	3	3	2	0	2	2	15	3
55	1	20	0	0	1	1	3	2	1	2	3	3	1	2	0	0	0	2	3	0	0	0	5	1
56	1	15	0	0	1	1	2	2	1	1	3	2	2	2	0	2	0	2	3	1	2	0	10	3
57	1	18	0	0	1	1	2	2	1	2	1	1	2	1	0	1	0	0	3	0	0	0	4	1
58	1	34	1	0	1	1	2	2	1	2	1	4	2	1	1	1	0	0	2	0	0	0	4	1
59	1	52	1	1	1	1	2	2	1	2	1	3	2	1	3	0	3	0	3	0	0	0	9	2
60	1	29	0	1	3	1	1	1	3	1	1	1	1	2	0	1	0	0	3	0	1	0	5	1
61	1	39	1	0	5	1	1	2	1	2	1	3	2	2	1	1	0	2	0	0	1	0	5	1
62	1	51	1	1	1	1	1	2	1	2	1	1	2	2	0	0	0	1	3	0	1	0	5	1
63	2	38	1	0	5	1	1	2	1	2	1	3	2	1	1	3	0	0	2	1	1	0	8	2
64	2	30	1	0	3	1	2	1	4	1	1	3	2	1	0	0	0	2	2	0	0	0	4	1
65	2	32	1	0	3	1	1	2	1	2	2	3	2	1	1	0	1	2	0	0	0	0	4	1
66	2	36	1	0	1	1	2	2	1	2	2	4	2	1	0	1	1	0	1	0	2	0	5	1
67	1	27	0	0	3	1	1	2	1	1	1	2	2	1	0	3	0	0	3	0	2	0	8	2
68	3	50	1	0	2	4	1	1	1	2	2	1	2	1	0	0	2	1	3	0	2	0	8	2
69	4	24	0	0	4	1	2	2	1	2	1	4	2	1	0	1	3	2	3	0	0	0	9	2
70	4	29	0	0	5	1	3	1	1	1	3	2	2	1	0	0	0	1	3	2	0	0	6	1
71	4	45	1	0	5	1	1	2	1	2	1	3	2	1	1	1	1	1	2	0	2	0	8	2
72	4	46	1	0	5	1	1	2	1	2	1	3	2	1	1	1	1	3	3	1	1	1	12	3
73	4	19	0	1	4	1	1	1	1	1	1	3	1	2	0	0	0	0	0	0	0	0	0	1
74	4	26	0	1	4	1	2	2	1	1	1	2	2	1	2	0	0	2	2	0	0	0	6	1
75	4	33	1	1	4	1	2	2	1	1	1	2	2	1	0	0	0	0	1	0	0	0	1	1
76	4	26	0	0	4	1	2	2	1	1	1	2	2	2	2	2	2	3	2	1	2	1	15	3
77	4	19	0	1	4	1	1	1	1	2	1	1	2	2	1	2	0	2	3	0	0	0	8	2
78	4	26	0	0	4	1	1	2	1	1	1	4	2	2	2	2	0	3	3	0	3	0	13	3
79	4	22	0	1	4	1	3	2	1	2	1	3	2	1	1	2	0	0	1	0	2	0	6	1
80	4	18	0	1	4	1	1	1	1	1	1	3	2	2	2	2	2	3	3	2	3	2	19	3
81	4	20	0	1	4	1	1	2	1	1	1	1	2	1	1	2	2	3	2	1	1	1	13	3
82	4	30	1	0	4	1	3	1	1	2	1	2	2	2	3	0	2	2	1	2	1	1	12	3
83	4	22	0	1	4	1	2	1	4	1	1	2	1	2	2	2	0	3	1	0	1	0	9	2
84	4	51	1	0	4	1	1	2	1	2	1	4	2	1	2	3	0	3	3	1	3	0	15	3
85	4	23	0	1	4	1	2	2	1	1	1	4	2	2	3	2	0	1	1	1	3	0	11	3
86	4	39	1	1	4	1	2	2	1	2	1	2	2	1	3	3	0	0	2	0	0	0	8	2
87	4	29	0	1	4	1	1	1	1	1	1	3	2	1	2	0	0	0	1	0	1	0	4	1
88	4	21	0	1	4	1	4	2	1	2	3	2	1	2	0	2	1	3	3	0	0	0	9	2
89	4	21	0	0	4	1	3	1	1	1	1	2	2	2	3	1	2	3	3	0	2	0	14	3
90	4	20	0	1	4	1	1	2	1	1	2	2	2	2	2	0	0	0	3	0	3	0	8	2
91	1	18	0	0	2	1	1	2	1	1	1	1	2	2	3	3	3	0	3	0	0	0	12	3

Apêndice B - continuação

Id	Loc	Idad	I_Rc	Sex	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	S1	S2	S3	S4	S5	S6	S7	S8	S_Tot	Dig
92	4	25	0	1	4	1	1	1	1	1	1	1	2	1	1	0	0	1	2	0	3	0	7	2
93	4	33	1	1	4	1	2	2	1	1	1	1	1	2	1	2	0	3	3	0	3	0	12	3
94	4	30	1	1	5	1	2	1	4	1	1	1	2	1	1	1	1	0	3	1	2	0	9	2
95	4	25	0	1	4	1	3	2	1	2	1	1	2	2	2	2	0	0	2	0	3	0	9	2
96	4	32	1	1	4	1	2	2	1	1	1	4	2	1	1	3	3	0	2	0	2	0	11	3
97	4	40	1	0	4	1	1	2	1	1	1	3	2	1	2	1	0	0	3	0	2	0	8	2
98	4	19	0	1	4	1	1	2	1	2	1	1	2	1	1	0	0	2	3	0	2	0	8	2
99	4	29	0	1	4	1	2	1	1	1	1	3	2	2	3	1	2	3	3	0	3	0	15	3
100	2	26	0	1	3	1	1	2	1	1	2	2	2	1	1	1	1	2	2	0	0	0	7	2