

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO  
PUC/SP**

**CORINA RODRIGUES**

**LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS:  
UM ESTUDO COMPARATIVO SOBRE O DESEMPENHO DE ALUNOS  
DE LICENCIATURA EM MATEMÁTICA, PEDAGOGIA E  
BACHARELADO EM ADMINISTRAÇÃO**

**MESTRADO PROFISSIONAL EM ENSINO DE MATEMÁTICA**

**São Paulo**

**2009**

PONTIFÍCIA UNIVERSIDADE CATÓLICA DE SÃO PAULO  
PUC/SP

CORINA RODRIGUES

LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS:  
UM ESTUDO COMPARATIVO SOBRE O DESEMPENHO DE ALUNOS  
DE LICENCIATURA EM MATEMÁTICA, PEDAGOGIA E  
BACHARELADO EM ADMINISTRAÇÃO

*Dissertação apresentada à Banca Examinadora da Pontifícia  
Universidade Católica de São Paulo, como exigência parcial para  
obtenção do título de **MESTRE PROFISSIONAL EM ENSINO DE  
MATEMÁTICA**, sob a orientação da **Profa. Dra. Sandra Maria  
Pinto Magina**.*

São Paulo

2009

**Banca Examinadora**

---

---

---

Autorizo, exclusivamente para fins acadêmicos e científicos, a reprodução total ou parcial desta Dissertação por processos de fotocopiadoras ou eletrônicos.

**Assinatura:** \_\_\_\_\_ **Local e Data:** \_\_\_\_\_

Para minha amada mãe Josefa, a quem devo minha vida

Para meu querido marido José Roberto

Para meu amado pai Aurélio (in memoriam)

E meu querido irmão Jefferson (in memoriam)

## AGRADECIMENTOS

---

Agradeço a Deus, pela sua infinita bondade em me conceder sabedoria e colocar as pessoas certas em meu caminho.

A minha família, pela compreensão da minha ausência em dias tão importantes. Em especial a minha Mãe e ao meu marido.

A minha sogra Annica, que tantas orações fez para que eu continuasse firme no propósito de realizar meu sonho.

À Professora Doutora Sandra Maria Pinto Magina, pela orientação, amizade, dedicação e paciência.

Ao grupo REPARE em Educação Matemática, pela imensa contribuição, pelos conselhos e docinhos; em especial ao Aparecido (Cido), Vera, Franciana (Fran), Aida, Irene e Eurivalda (Euri), por suas experiências e valiosas considerações.

Ao Doutor Armando Traldi Junior e Doutora Luzia Aparecida Palaro, que gentilmente aceitaram participar da Banca Examinadora e contribuir para a realização deste trabalho.

Aos Professores do Programa de Mestrado Profissional em Educação Matemática da PUC-SP, pelas contribuições para a minha formação.

Aos meus colegas de mestrado, pela irmandade, convívio, amizade, discussões e alegrias; em especial, Romeu, Sérgio, Fábio e Carlos.

A minha amiga Silvana, que hoje ocupa o lugar da irmã que nunca tive. Não tenho como agradecer tudo pelo que tem passado ao meu lado e pelas palavras sábias nas horas de fraqueza.

A Rose (inspetora) e aos meus amigos e Professores Cássia, Marcos, Edmar, Cida Couto, Tamura e Cícero, que sempre acreditaram e souberam me apoiar nas horas mais difíceis.

As minhas amigas e Professoras Renata, Dulce, Círcia e Adriana, que contribuíram para a conquista e realização deste trabalho. São todas muito importantes para mim!

Ao secretário Francisco, pela colaboração e atenção dispensadas nesse período.

À Diretoria de Ensino de São Bernardo do Campo, pela dedicação e paciência em preencher tantos papéis, em especial à Supervisora Helenir. Obrigada por tudo.

Por fim, à Secretaria de Estado de Educação, pela bolsa concedida e razão desta pesquisa.

Muitíssimo Obrigada!

A Autora

## RESUMO

---

A presente pesquisa investigou quais são os conhecimentos básicos de um grupo de alunos dos cursos de Licenciatura em Matemática, Pedagogia e bacharelado em Administração com relação à leitura e interpretação de gráficos e tabelas estudados na disciplina de Estatística. A escolha de investigar como a Estatística é ensinada na licenciatura em Matemática, dentre tantos cursos de licenciatura, deu-se por acreditarmos serem esses futuros professores de Matemática, os responsáveis por ensinar de uma forma mais analítica, a Estatística básica nas escolas. Os futuros Pedagogos, ao optarem pelo curso de Pedagogia, poderão introduzi-la nas escolas, porém de forma mais pictórica e menos analítica. O bacharelado em Administração foi escolhido por entendermos ser um curso que faz da Estatística básica uma “*disciplina de serviço*”, com ênfase na leitura e interpretação de gráficos e tabelas e pela aplicabilidade em sua carreira em contexto de usuário da Estatística. Tendo como hipótese que os alunos de Administração apresentariam melhor desempenho frente às situações que envolvem a leitura e interpretação de gráficos e tabelas do que os alunos de Licenciatura em Matemática e Pedagogia; aplicamos um teste diagnóstico em 174 sujeitos divididos da seguinte forma: 72 sujeitos de Licenciatura em Matemática, 48 sujeitos de Pedagogia e 54 de bacharelado em Administração. Quanto ao aspecto teórico-epistemológico nossa pesquisa segue uma abordagem empírico-analítica com perspectiva descritiva. O tipo de coleta segue os preceitos de uma pesquisa naturalista ou de campo e, a análise dos resultados, uma abordagem quali-quantitativa promovendo uma comparação entre os desempenhos dos três grupos. Essa análise nos evidenciou que os desempenhos dos sujeitos de Licenciatura em Matemática foram estatisticamente mais positivos que os desempenhos dos sujeitos de Administração e estes, mais positivos que os desempenhos dos sujeitos de Pedagogia, refutando nossa hipótese de pesquisa.

**Palavras-Chave:** Leitura e Interpretação de Gráficos e Tabelas; Registros de Representação Semiótica em Estatística; Educação Estatística; teste diagnóstico.

## ABSTRACT

---

This research investigated what are the basic skills of students in the mathematics, pedagogy and bachelor in Business Administration courses with respect to reading and interpreting of graphics and tables in the Statistics' subject. The choice to investigate how the statistics is taught in mathematics, among many graduate courses occurred because we believe that the future teachers of mathematics will be responsible for teaching the basic statistics of a analytical manner. So, when they will be educators, they will be responsible for introduction the same statistics in school, but in a more pictorial and less analytical manner. The Administration course was chosen because we believe that it is a course that makes of the basic statistics a "discipline of service", with emphasis on reading and interpreting of graphs and tables and the applicability of its career in contexts of user statistics. In hypothesis that students had better performance ahead situations that involving the reading and interpreting of graphs and tables than students of the Mathematics and Pedagogy courses, we apply a diagnostic test 174 students divided in the following way: 72 students in Mathematics course, 48 students in pedagogy course and 54 students in Administration course. About the theoretical and epistemological aspect, our research follows an empirical analytic approach with perspective descriptive. The type of collection follows the precepts of a naturalistic or field research, and the analysis of results, a qualitative and quantitative approach that organize comparison between the performances of three groups. This analysis showed us that in the performance of students in Mathematics course were statistically more positive than the performance of students in administration and these, more positive than the performance of the students in pedagogy, refuting our research hypothesis.

**Keywords:** Reading and Interpretation of Graphs and Tables; records of semiotic representation in Statistics; Statistics education; diagnostic test.

## SUMÁRIO

APRESENTAÇÃO – Meus caminhos.....	15
<b>CAPÍTULO I.....</b>	<b>16</b>
1.1. INTRODUÇÃO.....	16
1.2. PROBLEMÁTICA.....	18
1.3. OBJETIVO E QUESTÃO DE PESQUISA.....	25
1.4. DESCRIÇÃO RESUMIDA DA DISSERTAÇÃO.....	30
<b>CAPÍTULO II.....</b>	<b>32</b>
2.1. MEDIDAS DE TENDÊNCIA CENTRAL: COMPREENDENDO SEUS CONCEITOS BÁSICOS.....	33
2.1.1. MODA.....	34
2.1.2. A MEDIANA.....	36
2.1.3. A MÉDIA.....	37
2.1.4. MODA, MEDIANA OU MÉDIA.....	38
2.2. O PERCURSO TEÓRICO DA EDUCAÇÃO ESTATÍSTICA.....	40
2.2.1. EDUCAÇÃO ESTATÍSTICA: UM DESAFIO.....	41
2.2.2. O LETRAMENTO ESTATÍSTICO E O RACIOCÍNIO ESTATÍSTICO.....	43
2.2.3. A CONSTRUÇÃO, LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS.....	47
2.2.3.1. CONSTRUÇÃO DE TABELAS.....	49
2.2.3.2. LEITURA E INTERPRETAÇÃO DE TABELAS.....	52
2.2.3.3. CONSTRUÇÃO DE GRÁFICOS.....	53
2.2.3.4. LEITURA E INTERPRETAÇÃO DE GRÁFICOS.....	57
2.2.4. REGISTROS DE REPRESENTAÇÃO SEMIÓTICA: A TEORIA DE RAYMOND DUVAL.....	61
<b>CAPÍTULO III.....</b>	<b>71</b>
3.1. A TEORIA METODOLÓGICA: PASSO A PASSO.....	71
3.2. O DELINEAMENTO DO UNIVERSO DE ESTUDO.....	74
3.2.1. OS SUJEITOS.....	74
3.2.2. O MATERIAL DE ESTUDO.....	77
3.2.2.1. O TESTE: SEUS OBJETIVOS E CONSTRUÇÃO.....	78
3.2.2.2. UMA ANÁLISE A PRIORI DO TESTE.....	80
<b>CAPÍTULO IV.....</b>	<b>94</b>
4.1. ANÁLISE QUANTITATIVA E QUALITATIVA: O DESEMPENHO TRADUZIDO EM NÚMEROS.....	94
4.1.1. A ANÁLISE GERAL DOS GRUPOS.....	96
4.1.2. A ANÁLISE QUANTITATIVA DO DESEMPENHO POR UNIDADE DE ANÁLISE.....	100
4.1.2.1. LEITURA E INTERPRETAÇÃO DE GRÁFICOS E SUAS CATEGORIAS.....	102
4.1.2.2. LEITURA E INTERPRETAÇÃO DE TABELAS E SUAS CATEGORIAS.....	116
4.1.2.3. REPRESENTAÇÃO SEMIÓTICA EM ESTATÍSTICA E SUAS CATEGORIAS.....	126
4.1.2.4. O DESEMPENHO EM CÁLCULO DE MODA, MÉDIA E MEDIANA.....	145

<b>CAPÍTULO V .....</b>	<b>148</b>
5.1. RESGATE DO PERCURSO DA PESQUISA .....	148
5.2. SÍNTESE DOS PRINCIPAIS RESULTADOS OBTIDOS .....	150
5.2.1. SOBRE O DESEMPENHO GERAL NO TESTE .....	150
5.2.2. SOBRE O DESEMPENHO NAS UNIDADES DE ANÁLISE .....	151
5.3. RESPONDENDO A QUESTÃO DE PESQUISA.....	154
5.4. Sugestão para futuras pesquisas.....	159
<b>REFERÊNCIAS .....</b>	<b>162</b>

## LISTA DE FIGURAS

Figura 1: Esquema de alguns diferentes usos da Estatística. ....	27
Figura 2: Representação gráfica de distribuições unimodal e bimodal .....	35
Figura 3: Exemplo de uma distribuição simétrica, unimodal. ....	39
Figura 4: Exemplos das posições relativas das medidas de tendência central em uma distribuição .....	40
Figura 5: Esquema do pensamento em relação ao raciocínio estatístico e matemático.....	47
Figura 6: Exemplo da estrutura de uma tabela simples, retirada de LEVIN e FOX, 2007, p. 28. ....	50
Figura 7: Exemplo da estrutura de uma tabela de dupla entrada, retirada de LEVIN e FOX, 2007, p.29.....	51
Figura 8: Exemplo da estrutura da construção de um gráfico de colunas. ....	54
Figura 9: Exemplo da estrutura da construção de um gráfico de barras.....	55
Figura 10: Exemplo de gráfico de colunas promovendo comparação. ....	55
Figura 11: Exemplo de gráfico de barras promovendo a comparação.....	56
Figura 12: Exemplo da estrutura de um gráfico de setor. ....	56
Figura 13: Estado civil de 215 pessoas da Cidade de Vistajoia em fev 2000 .....	59
Figura 14: Gráfico de setor com estado civil de 215 pessoas da Cidade de Vistajoia em fev 2000. ....	59
Figura 15: Estado civil de 215 pessoas da Cidade de Vistajoia em fev 2000 .....	60
Figura 16: Gráfico de setor sobre o gosto pela Matemática de 50 alunos .....	62
Figura 17: Gráfico de colunas sobre o gosto pela Matemática de 50 alunos .....	63
Figura 18: Mudança da escala do gráfico da variável idade .....	66
Figura 19: Gráfico de setor da distribuição das disciplinas favoritas de 10 alunos. ....	67
Figura 20: Gráfico cartesiano da função $f(x) = x+3$ .....	68
Figura 21: Demonstração de uma conversão congruente .....	69
Figura 22: Demonstração de uma conversão não-congruente .....	70
Figura 23: Desenho dos sujeitos da pesquisa .....	76
Figura 24: Desenho da pesquisa .....	77
Figura 25: Desenho dos objetivos de cada questão. ....	79
Figura 26: Desempenho total dos grupos no teste e tabela com respectivo teste estatístico de Tukey. ....	97
Figura 27: Exemplo de resposta de sujeitos de Pedagogia para os itens de perfil 1.15 e 1.16 .....	98
Figura 28: Boxplot com o desempenho dos grupos no teste. ....	99
Figura 29: : Exemplo de respostas dos sujeitos de Pedagogia quanto ao tempo de estudo.....	100
Figura 30: Desempenho total dos grupos em cada Unidade de Análise. ....	101
Figura 31: Desempenho dos sujeitos, segundo os grupos, nas categorias de Leitura e Interpretação de Gráficos e tabela com respectivos testes de Tukey .....	103
Figura 32: Exemplo dos gráficos utilizados na questão Q.2.1. ....	105
Figura 33: Gráfico de colunas com a Taxa de respostas erradas à questão Q.2.1. ....	106
Figura 34: Exemplo de respostas dos sujeitos para o item Q.3.a.b. ....	108
Figura 35: Exemplo de respostas dos sujeitos ao item Q.6.2. ....	110
Figura 36: Exemplo de resposta de um sujeito do GFP ao item Q.3.b. ....	112
Figura 37: Exemplo de resposta dos sujeitos do GFADM ao item Q.3.b. ....	112
Figura 38: Exemplos de respostas dos sujeitos do GFLM e GFADM ao item Q.6.3.....	113
Figura 39: Desempenho dos sujeitos, segundo os grupos, no item Q.6.1 e tabela do teste de Tukey.....	114
Figura 40: Exemplo de resposta de um sujeito do GFLM para o item Q.6.1. ....	115
Figura 41: Desempenho dos sujeitos, segundo os grupos, nas categorias de Leitura e Interpretação de Tabelas e tabela com respectivos testes de Tukey. ....	116
Figura 42: Exemplos de respostas dos sujeitos ao item Q.5.2. ....	118

Figura 43: Exemplos de respostas dos sujeitos no item Q.4.1.a e Q.4.1.b.....	122
Figura 44: Exemplos de respostas dos sujeitos ao item Q.4.1.a e Q.4.1.b.....	122
Figura 45: Exemplos de respostas dos sujeitos ao item Q.4.2. ....	123
Figura 46: Exemplo de resposta de um sujeito do GFLM ao item Q.5.1.....	125
Figura 47: Desempenho dos grupos na unidade de análise Representação Semiótica em Estatística.....	127
Figura 48: Exemplo de resposta de um sujeito do GFLM ao item Q.7.....	130
Figura 49: Exemplo de resposta de um sujeito do GFLM ao item Q.7.....	130
Figura 50: Exemplo de resposta de um sujeito do GFADM ao item Q.7.....	130
Figura 51: : Exemplo de resposta de um sujeito do GFADM ao item Q.7.....	131
Figura 52: Exemplos de respostas de dois sujeitos do GFP ao item Q.10.a. ....	132
Figura 53: Exemplos de respostas de dois sujeitos do GFLM ao item Q.10.a.....	132
Figura 54: Exemplos de respostas de dois sujeitos do GFLM ao item Q.4.3.....	134
Figura 55: Exemplos de respostas dos sujeitos do GFADM à questão Q.4.3. ....	135
Figura 56: Exemplo de resposta de um sujeito do GFLM à questão Q.4.3.....	135
Figura 57: Exemplo de resposta de um sujeito do GFADM à questão Q.4.3.....	136
Figura 58: Exemplos de respostas dos sujeitos do GFLM e GFADM, respectivamente ao item Q.8. ....	137
Figura 59: Exemplos de respostas dos sujeitos do GFADM e GFLM à questão Q.8, respectivamente. ....	138
Figura 60: Exemplos de respostas dos sujeitos do GFADM e GFLM à questão Q.8, respectivamente ....	138
Figura 61: Exemplo de resposta de um sujeito do GFP para a questão Q.10.b. ....	141
Figura 62: Exemplos de respostas de dois sujeitos do GFLM ao item Q.10.b.....	141
Figura 63: Exemplo de resposta de um sujeito do GFLM ao item Q.4.4.....	144
Figura 64: Exemplo de resposta de um sujeito do GFADM ao item Q.4.4.....	145
Figura 65: Exemplo de resposta de um sujeito do GFP ao item Q.4.4. ....	145
Figura 66: Desempenho dos grupos em cálculos de moda, média, mediana ....	146
Figura B.67: Exemplo de variáveis qualitativas nominais, retirada de LEVIN e FOX, 2007, p. 9.....	174
Figura B.68: Exemplo de variáveis qualitativas ordinais, retirada de LEVIN e FOX, 2007, p. 11. ....	174
Figura B.69: Exemplo de nível intervalar, retirada de LEVIN e FOX, 2007, p. 12.....	175
Figura B.70: Esquema de classificação de variáveis. ....	176
Figura B.71: Gráfico de Setores da distribuição de 200 alunos ....	180
Figura 72: Gráfico de Colunas da distribuição de 200 alunos.....	180
Figura B.73: Gosto por Matemática e Língua Portuguesa, baseado em CAZORLA e SANTANA, 2006, p. 18...	181
Figura 74: Histograma, baseado em NOVAES e COUTINHO, 2008, p. 26.....	183
Figura B.75: Polígono de Freqüências e Histograma, baseado em NOVAES e COUTINHO, 2008, p. 27. ....	183
Figura B.76: Polígono de Freqüências, baseado em NOVAES e COUTINHO, 2008, p. 27. ....	184
Figura B.77: Diagrama de ramo e folha das notas do 2º bimestre do ano de 2001 de 25 alunos da 7ª A ....	187
Figura C.78: Gráfico demonstrativo do cálculo da Moda em dados agrupados em intervalos. ....	188
Figura C.79: Histograma da distribuição do tempo de vida útil da peça XY. ....	189
Figura E.80: Distribuição dos tempos (minutos) que 16 entrevistados passaram ao telefone. ....	203

## LISTA DE TABELAS

Tabela 1: Distribuição de 1550 peças A, .....	35
Tabela 2: Valor da diária escolhida pelos hóspedes da cidade Vistajóia.....	36
Tabela 3: Distribuição de frequência do gosto pela Matemática de 50 alunos .....	62
Tabela 4: Disciplinas favoritas de 10 alunos da escola Ratatui .....	65
Tabela 5: Distribuição das disciplinas favoritas de 10 alunos .....	65
Tabela 6: Distribuição das disciplinas favoritas de 10 alunos .....	66
Tabela 7: Desempenho dos sujeitos nas questões de Leitura e Interpretação de Gráficos .....	107
Tabela 8: Desempenho dos sujeitos nos itens de Leitura e Interpretação de Gráficos .....	111
Tabela 9: Desempenho dos sujeitos, segundo os grupos, no item Q.5.2 no “Nível Básico”.....	118
Tabela 10: Desempenho dos sujeitos, segundo os grupos, nas questões no “Nível Intermediário” .....	119
Tabela 11: Desempenho dos sujeitos, segundo os grupos, no item Q.4.1.a.....	120
Tabela 12: Desempenho dos sujeitos, segundo os grupos, no item Q.4.1.b. ....	121
Tabela 13: Desempenho dos grupos no item Q.4.2 no “Nível Intermediário”. ....	123
Tabela 14: Desempenho dos sujeitos, segundo os grupos, nas questões no “Nível Avançado”.....	124
Tabela 15: Desempenho dos sujeitos na “conversão do registro gráfico para o registro tabular”. ....	128
Tabela 16: Desempenho dos sujeitos, segundo os grupos na questão Q.7 com respectivo teste de Tukey.....	129
Tabela 17: Desempenho dos sujeitos, segundo os grupos no item Q.10.a com respectivo teste de Tukey.....	131
Tabela 18: Desempenho dos sujeitos na “conversão do registro tabular para o registro gráfico”. ....	133
Tabela 19: Desempenho dos sujeitos, segundo os grupos no item Q.4.3 com respectivo teste de Tukey.....	134
Tabela 20: Desempenho dos sujeitos no item Q.8 e tabela com respectivo teste de Tukey. ....	136
Tabela 21: Desempenho dos sujeitos na “conversão do registro tabular para o registro numérico”.....	139
Tabela 22: Tabela com respectivo teste de Tukey e desempenho dos sujeitos no item Q.5.2.....	140
Tabela 23: Tabela com respectivo teste de Tukey e desempenho dos sujeitos no item Q.5.2.....	140
Tabela 24: Tabela com respectivo teste de Tukey e desempenho dos sujeitos no item Q.6.3.....	142
Tabela 25: Tabela com respectivo teste de Tukey e desempenho dos sujeitos no item Q.4.4.....	143
Tabela B.26: Gosto pela Matemática e por Língua Portuguesa de 200 alunos matriculados na 7ª série .....	178
Tabela B.27: Distribuição de 200 alunos quanto ao gosto pela Matemática.....	179
Tabela B.28: Distribuição de 200 alunos quanto ao gosto pela Língua Portuguesa. ....	179
Tabela B.29: Relação entre gosto pela Matemática e gosto por Língua Portuguesa .....	181
Tabela B.30: Distribuição de frequências das idades de 153 pessoas que compraram livros .....	182
Tabela B.31: Distribuição de 153 pessoas segundo a faixa etária.....	183
Tabela B.32: Distribuição de 153 pessoas segundo a idade, expressa em anos. ....	186
Tabela B.33: Distribuição de frequência das notas no 2º bimestre do ano de 2001.....	187
Tabela C.34: Distribuição do tempo de vida útil da peça XY .....	189
Tabela D.35: Nº de equipamentos distribuídos segundo o tempo (em meses) . ....	192
Tabela D.36: Nº de equipamentos distribuídos segundo o tempo (em meses) . ....	193
Tabela D.37: Nº de equipamentos distribuídos segundo o tempo (em meses) . ....	193
Tabela D.38: Nº de equipamentos distribuídos segundo o tempo (em meses) .....	194
Tabela D.39: Distribuição das pessoas que usam um determinado tipo de celular segundo as idades (anos)...	195
Tabela E.40: QI de oito entrevistados, retirado de LEVIN e FOX, 2007, p. 82. ....	199
Tabela E.41: Desvios de um conjunto de variáveis em torno da média $X$ , .....	200
Tabela E.42: Distribuição da diária de pousadas escolhidas pelos hóspedes.....	201
Tabela E.43: Distribuição das pessoas que usam um determinado tipo de celular segundo as idades (anos),..	202

## LISTA DE QUADROS

Quadro 1: Classificação dos níveis de usuários da Estatística, retirado de CAZORLA, 2002, p. 24 .....	22
Quadro 2: Demonstração da posição de Q2, adaptado de NOVAES e COUTINHO, 2008, p. 32. ....	37
Quadro 3: Generalização da distribuição de uma tabela simples .....	50
Quadro 4: Classificação dos diferentes registros mobilizáveis no funcionamento matemático de Duval .....	64
Quadro 5: Distribuição dos itens em relação às unidades de análise e suas categorias.....	95
Quadro B.6: Fórmula da Regra de Sturges e do Critério da raiz. ....	184
Quadro B.7: Fórmula para calcular a amplitude dos intervalos em distribuição de freqüências agrupadas. ....	185
Quadro B.8: Quadro das notas do 2º bimestre do ano de 2001 de 25 alunos da 7ª A da escola Ratatui.....	186
Quadro C.9: Demonstração do cálculo de Moda por meio de Semelhança de Triângulos e Fórmula de Czuber.....	188
Quadro C.10: Cálculo da moda da distribuição de dados intervalares do tempo de vida útil da peça XY .....	190
Quadro D.11: Demonstração de determinação dos quartis .....	191
Quadro D.12: Demonstração de determinação dos quartis, retirado de NOVAES e COUTINHO, 2008, p. 33 ...	192
Quadro D.13: Demonstração de determinação dos quartis, baseado em NOVAES e COUTINHO, 2008, p. 34.	193
Quadro D.14: Demonstração de determinação dos quartis, baseado em NOVAES e COUTINHO, 2008, p. 35.	194
Quadro D.15: Demonstração de determinação dos quartis, baseado em NOVAES e COUTINHO, 2008, p. 36.	195
Quadro D.16: Esquema da determinação dos quartis, retirado de NOVAES e COUTINHO, 2008, p. 36. ....	195
Quadro D.17: Demonstração de determinação dos quartis, adaptado de NOVAES e COUTINHO, 2008, p. 37.	196
Quadro D.18: Demonstração de determinação dos quartis, retirado de NOVAES e COUTINHO, 2008, p. 37. ...	196
Quadro D.19: Demonstração de determinação dos quartis, retirado de NOVAES e COUTINHO, 2008, p. 38. ...	197
Quadro D.20: Esquema da determinação dos quartis da distribuição das pessoas. ....	197
Quadro E.21: Analogia da alavanca e fulcro para a média, retirado de LEVIN e FOX, 2007, p. 83. ....	200
Quadro E.22: Demonstração das fórmulas para cálculos de média aritmética de dados agrupados .....	201

## LISTA DE APÊNDICES

APÊNCIDE A - O Instrumento Diagnóstico.....	166
APÊNCIDE B - Medidas de tendência central .....	171
APÊNCIDE C - A Moda .....	188
APÊNCIDE D - A Mediana.....	191
APÊNCIDE E - A Média .....	199
APÊNCIDE F - As medidas de variabilidade .....	205

Sou professora de Matemática há dezessete anos, sempre atuando na cidade de São Bernardo do Campo, localizada na região do Grande ABC paulista. Durante esse período nunca parei de estudar e procurei sempre a atualização profissional, por meio de cursos ligados ao ensino, acreditando ser esse o caminho para o desenvolvimento.

Quando da realização do concurso público na rede estadual de ensino em 2004, tornei-me titular de cargo efetivo na disciplina de Matemática.

Durante muito tempo ministrei aulas somente para o Ensino Fundamental, no entanto, ultimamente trabalho também com o Ensino Médio.

Quando a Diretoria de Ensino da minha região, convocava os professores para cursos voltados ao ensino da Matemática, sempre ficava na expectativa de acrescentar novos conhecimentos aos anteriores e prontamente fazia minha inscrição.

Esta busca por atualização profissional direcionou-me a vários caminhos, até chegar à pós-graduação no nível de especialização em Educação Matemática da Pontifícia Universidade Católica – PUC-SP. Tal especialização foi oferecida aos professores da rede pública Estadual e, ainda cursando, percebi que este era meu caminho.

Portanto, quando houve a oportunidade de fazer Mestrado na área do Ensino da Matemática, não tive dúvidas, fiz minha inscrição no processo de seleção, uma vez aprovada iniciei o curso.

Meus estudos foram custeados pela Secretaria da Educação do Estado de São Paulo, a quem devo parte desta realização.

### 1.1. INTRODUÇÃO

Este trabalho foi pensado e idealizado na linha de pesquisa “A Matemática na Estrutura Curricular e Formação de Professores”, a partir de estudos realizados no Programa de Estudos Pós-graduados em Educação Matemática da Pontifícia Universidade Católica de São Paulo (PUC-SP) nos anos de 2007 e 2008.

Sob a orientação da Professora Doutora Sandra Maria Pinto Magina, esta pesquisadora ingressou no grupo REPARE em Educação Matemática (Reflexão – planejamento – ação - reflexão). Dentro deste grupo, temos um subgrupo que pesquisa assuntos referentes à Educação Estatística do Ensino Fundamental ao Superior.

No momento do ingresso desta pesquisadora neste subgrupo, podíamos encontrar pesquisas direcionadas ao processo ensino-aprendizagem da Estatística no Ensino Fundamental I, Ensino Fundamental II e no Ensino Médio, bem como pesquisas sobre os conhecimentos estatísticos dos professores que ministram suas aulas nestes níveis. Faltava, porém, uma pesquisa direcionada ao Nível Superior, para que pudéssemos abarcar todos os níveis de ensino.

Esta pesquisa começou a tomar corpo e direcionar-se para a Educação Estatística por dois motivos: o primeiro é que a pesquisadora já apresentava interesse em trabalhos sobre o ensino da Estatística. Tal área do conhecimento é uma ferramenta útil e serve não somente à Matemática, mas aos campos da saúde, política, economia, desporto, ciência e cada vez mais está presente no cotidiano dos alunos.

O segundo é que para o ciclo de pesquisas do subgrupo se fechar em torno de todos os níveis de ensino faltava uma pesquisa com o Ensino Superior. Portanto, com o ingresso desta pesquisadora no grupo REPARE a questão foi sanada e, finalmente, foi possível abraçar a investigação acerca da leitura e interpretação de gráficos e tabelas no Ensino Superior.

Com relação às pesquisas neste subgrupo estatístico podemos encontrar o trabalho de Veras (em andamento), que visa a identificar e analisar os conhecimentos dos professores do Ensino Fundamental I sobre leitura e interpretação de gráficos apresentados nos meios de comunicação.

Já o trabalho de Pereira (em andamento) pretende investigar os avanços e limitações de uma intervenção de ensino, pautada na leitura e interpretação de gráficos e tabelas, aplicada em uma 5ª série (6º ano) do Ensino Fundamental I.

Contamos com a pesquisa de Pagan (em andamento), que se baseia na Proposta Curricular do Estado de São Paulo (2009) e investiga sobre qual educador está mais preparado para trabalhar com Estatística em sala de aula: o professor de Matemática, que possui um arcabouço matemático importante na Estatística, ou o de Geografia, que possui os conhecimentos geográficos e desenvolve o perfil de um usuário estatístico.

É preciso mencionar, ainda, o trabalho de Leite (em andamento). Sua pesquisa versa sobre que contribuições uma intervenção de ensino, pautada na estimativa dos cálculos de moda, média e mediana, pode oferecer aos alunos do Ensino Médio. O enfoque dado à análise de Leite se refere à leitura e interpretação de gráficos e tabelas.

A pesquisa de Reis (em andamento) pretende analisar como o ENEM (Exame Nacional do Ensino Médio) aborda gráficos e tabelas em suas provas. A análise abrange, sobretudo, o desempenho de um grupo de alunos do Ensino Médio em tais questões.

Nesse contexto, o presente trabalho propõe uma reflexão sobre os conceitos estatísticos básicos dos alunos de Licenciatura em Matemática, Pedagogia e bacharelado em Administração. Pretende-se, para tanto, fazer um estudo

comparativo entre os conhecimentos dos alunos dos três cursos em relação à leitura e interpretação adequada de gráficos e tabelas.

## 1.2. PROBLEMÁTICA

Nas últimas décadas percebemos que as pessoas convivem em uma sociedade dinâmica e rica em informações. Uma boa parte destas informações nos chega por meio de vários tipos de textos e/ou imagens, muitos dos quais trazem em seu conteúdo a Estatística como forma de comunicação ou explicação.

Acreditamos que muitas dessas pessoas frequentam ou frequentaram a escola ou a universidade. Logo, por serem estas responsáveis pela educação formal de seus alunos, incluindo o ensino da Estatística, poderiam também fazer com que lessem e interpretassem com cuidado e responsabilidade as formas de representação de dados estatísticos, proporcionando a eles maior participação crítica na sociedade.

Nesse sentido, o ensino da Estatística se faz necessário não só pelo seu fim em si mesmo, com ênfase em seus conceitos e métodos próprios, mas para fazer com que o sujeito seja crítico frente a uma informação. Isso é necessário para ele tomar decisões com base na informação analisada em vez de realizar cálculos estatísticos e ler os dados contidos em gráficos de modo automático.

Para Kirk, Eggen e Kauchak (1980, apud CURCIO, 1989), mesmo que os alunos tenham habilidade de ler os dados contidos numa figura gráfica, o que é muito importante, ainda é necessário que eles desenvolvam um olhar para além dos números apresentados, fazendo, assim, uma análise exploratória dos dados.

A análise exploratória dos dados é, para Batanero, Estepa e Godino (1991), uma filosofia que consiste

no estudo dos dados a partir de todas as perspectivas e com todas as ferramentas possíveis, incluindo as já existentes. O propósito é extrair toda a informação possível, gerar novas hipóteses no sentido de construir conjecturas sobre as observações que dispomos. (BATANERO; ESTEPA; GODINO, 1991, p.2)

Acreditamos, assim, que há necessidade da discussão de como as informações Estatísticas chegam às pessoas e, ainda, de como estas realizam a análise e a interpretação dos dados e de que maneira elaboram conjecturas.

Uma das formas de transmissão de informações à população se dá por tabelas e gráficos diversificados. Percebemos que ambas as formas de representar e apresentar dados estatísticos estão muito em voga na mídia. Podemos constatar esse fato ao abrirmos jornais e revistas que trazem todo o tipo de informação: políticas, econômicas, meteorológicas, questões de cunho social. Também a televisão, um poderoso veículo de comunicação em massa, faz uso constante de tais recursos. A internet passou a contribuir com o infográfico<sup>1</sup>, tendo este alcançado maior destaque entre os principais recursos utilizados pela mídia impressa.

Esses infográficos destacam-se, sobretudo, pela *“predominância da linguagem visual, a conectividade entre texto e imagem e a clareza no tratamento da informação, oferecendo ao público uma noção mais rápida e eficaz dos sujeitos, do tempo e do espaço da notícia.”* (MÓDOLO e GOUVEIA JUNIOR, 2007, p.1). Acreditamos que esse tipo de comunicação, o infográfico, também faz parte do universo informativo acessado pelo cidadão e por isso não deve ficar de fora tanto da Educação Estatística quanto da discussão sobre como se estão fazendo a leitura de gráficos e tabelas.

Com relação à leitura e compreensão dos dados gráficos, Frances R. Curcio (1989) define três níveis de compreensão gráfica: 1º nível – a leitura dos dados; 2º nível – a leitura entre os dados e pressupõe, como 3º nível, a leitura além dos dados que serão discutidas no Capítulo II. Em outros trabalhos de Curcio, posteriores aos de 1989, aparece um quarto nível de compreensão dos dados; no entanto, para essa pesquisa assumiremos somente os três níveis de compreensão gráfica de 1989.

Para a leitura e interpretação de tabelas Howard Wainer (1992) desenvolveu uma estrutura teórica para gráficos a partir da revisão do trabalho de Bertin (1973), deixando claro, porém, que esta mesma estrutura pode ser empregada quando da análise de representações tabulares. Os níveis de leitura podem ser classificados como básico, intermediário e avançado, e serão discutidos no Capítulo II.

---

<sup>1</sup> Segundo o dicionário Aurélio, infografia é a técnica de combinar desenhos, fotos, gráficos, etc. para a apresentação dramatizada de dados. §Infográfico adj.

Podemos então acreditar que há crescente necessidade de munir nossos alunos cidadãos com ferramentas para que desenvolvam leitura e compreensão dos dados gráficos até o 3º nível, ou pelo menos, até o 2º nível de Curcio (1989). Também é esperável que façam a leitura de dados tabulares até o nível avançado ou pelo menos, ao intermediário de Wainer (1992).

Pensamos que este papel, o de fazer com que os alunos desenvolvam os três níveis de compreensão gráfica de Curcio (1989) e os três níveis de leitura tabular de Wainer (1992), poderia ser desenvolvido por professores do Ensino Fundamental ao Ensino Superior. Dessa forma, haveria maior contribuição para o desenvolvimento do pensamento estatístico ao longo dos anos escolares.

Quando voltamos nosso olhar para o ensino da Estatística no Ensino Fundamental, encontramos nos Parâmetros Curriculares Nacionais (PCN) de 1998 a Estatística como um bloco de conteúdo denominado Tratamento da Informação, a saber:

Integrarão este bloco estudos relativos a noções de Estatística e de probabilidade, além dos problemas de contagem que envolvem o princípio multiplicativo. Evidentemente, o que se pretende não é o desenvolvimento de um trabalho baseado na definição de termos ou de fórmulas envolvendo tais assuntos. (BRASIL, 1998, p.52).

Já no documento Orientações Complementares aos Parâmetros Curriculares Nacionais (PCN+) de 2002, a Estatística é tratada como um tema estruturador denominado “Análise de Dados”:

A análise de dados (...) propõe-se que constitua o terceiro eixo ou tema estruturador do ensino, e tem como objetos de estudo os conjuntos finitos de dados, que podem ser numéricos ou informações qualitativas, o que dá origem a procedimentos bem distintos daqueles dos demais temas, pela maneira como são feitas as quantificações, usando-se processos de contagem combinatórios, frequências e medidas estatísticas e probabilidades. Este tema pode ser organizado em três unidades temáticas: Estatística, Contagem e Probabilidade. (BRASIL, 2002, p.126).

Isto posto, podemos concluir que nos cursos superiores voltados à Licenciatura, como Pedagogia ou Licenciatura em Matemática, seria desejável que os futuros professores dominassem minimamente a Estatística básica por ela fazer parte do bloco de conteúdo ou tema estruturador a ser ensinado. É preciso ressaltar, ainda, que cada vez mais frequente encontramos capítulos dedicados à Estatística nos livros didáticos. Isso justifica, portanto, a importância de refletir sobre como a Estatística é ensinada no Ensino Superior.

Com relação à atenção que as instituições de Ensino Superior têm dispensado ao curso de Estatística, parece-nos que algumas destas podem não estar cumprindo seu papel no desenvolvimento do pensamento estatístico.

De fato, em minha formação nos tempos de graduação (1998 a 2000), foi possível comprovar que a disciplina Estatística, naquela turma, era tratada apenas como um amontoado de fórmulas que precisavam ser decoradas, o que não favorecia o verdadeiro sentido dos conteúdos ensinados.

A Estatística poderia ter sido apresentada como uma ferramenta importante para a leitura de mundo, durante a minha graduação. Nem sempre os alunos eram estimulados a pensar em todo o potencial oferecido pela disciplina; somente se cumpria o rol de exercícios propostos sem realizar a reflexão e a análise exploratória dos dados oferecidos.

Pensamos que o aprendizado voltado para a análise exploratória de dados talvez seja uma prática mais desenvolvida nas salas de aula nos cursos de nível Superior que formarão Estatísticos, enquanto outros cursos, que utilizarão a Estatística em um contexto de usuário, podem não se aprofundar tanto no assunto.

A Estatística se faz presente, ainda, em cursos superiores que almejem fazer com que seus alunos sejam usuários da Estatística, como os de Pedagogia, Administração, Ciências Sociais, entre outros. Quando a disciplina Estatística é ensinada com o caráter de aplicabilidade, Wada (1996) a classifica como uma “*disciplina de serviço*”.

As graduações que oferecem Estatística como uma “*disciplina de serviço*” costumam mantê-la na grade curricular apenas uma disciplina introdutória, voltada para noções de Estatística básica, análise exploratória dos dados, de inferência e de probabilidade, restringindo tais conhecimentos à especificidade do curso.

Deste modo, Cazorla (2002), baseada em literaturas sobre o ensino da Estatística em um contexto de usuário, elaborou um quadro em que a classificação dos níveis deste usuário aparece conforme seu nível de instrução:

Nível	Características	Conceitos/procedimentos	Nível de instrução
I	Consumidor de informações veiculadas pela mídia	Interpretação de tabelas e gráficos, medidas de tendência central e dispersão. Noções de probabilidade	Ensino Fundamental e Médio
II	Consumidor/produtor de relatórios de levantamentos de dados	Método científico, amostragem, análise exploratória de dados	Ensino Superior: Estatística Básica
III	Consumidor/produtor de relatórios estatísticos	Inferência estatística: Estimação de parâmetros e teste de hipóteses	Ensino Superior: Inferência estatística
IV	Consumidor/produtor de relatórios estatísticos complexos	Base sólida em Estatística, ao nível de usuário	Ensino Superior: Estatística avançada
V	Consumidor/produtor de relatórios estatísticos complexos – Consultor	Base sólida em Estatística ao nível de graduação	Ensino Superior: Bacharel em Estatística
VI	Consumidor/produtor de técnicas estatísticas	Base sólida em Estatística Matemática	Ensino Superior: Pós-graduação em Estatística

Quadro 1: Classificação dos níveis de usuários da Estatística, retirado de CAZORLA (2002, p. 24)

Podemos observar no Quadro 1 que a Estatística se consolida como uma ferramenta de trabalho nos cursos superiores independentemente da área escolhida pelo aluno, uma vez que encontramos quatro dos cinco níveis destinados à graduação.

O cenário da educação Estatística no Ensino Superior brasileiro é no mínimo alarmante: a falta de interesse por parte dos alunos que cursam esta disciplina é costumeiramente apontada como um dos itens de preocupação. Encontramos no trabalho de Grácio, Oliveira e Oliveira (s.d) a sinalização da falta de motivação dos alunos da “*disciplina de serviço*” de Biblioteconomia, Ciências Sociais, Fonoaudiologia e Pedagogia. Isso ocorre devido à falta de visão da aplicabilidade da Estatística no exercício das profissões.

Por outro lado, deparamo-nos com a pesquisa de Wada (1996), que analisou o discurso de alguns professores universitários das “*disciplinas de serviço*” com relação à atitude de seus alunos com a Estatística. Nela, o autor observa que os professores entrevistados denunciam “*a falta de interesse dos alunos e até mesmo o desconhecimento quanto ao seu uso*”. Acusam como fator complicador a inadequação do momento dos cursos em que ela é oferecida, uma vez que a disciplina é ministrada no início da graduação; ainda, nota-se a provável imaturidade

nos alunos – “o aluno faz a estatística como uma disciplina obrigatória...não sabem nem prá que vai usar aquilo... Daí, quando ele chegar no estágio... que ele descobre a importância... (sic)[C-7]”, (WADA, 1996, p.148).

Mas qual a importância da Estatística na formação profissional no Ensino Superior? Quando a Estatística assume o papel de conteúdo de uma “*disciplina de serviço*”, podemos dizer que sua importância está na orientação para cuidados necessários quanto à metodologia de coleta de dados, a análise exploratória desses dados e quais as melhores formas de representar os resultados alcançados. No entanto, a finalidade e a importância da Estatística como “*disciplina de serviço*” ainda são questões muito discutidas na comunidade científica, principalmente no quesito quais as melhores estratégias de ensino.

Outro ponto importante a ser discutido é a “supervalorização” da Matemática dentro da Estatística como “*disciplina de serviço*”. Silva et al (2002), baseada em literatura da Educação Matemática, argumenta que:

Uma questão muito discutida é como utilizar adequadamente a matemática nas disciplinas de estatística. Salienta-se a importância de reforçar o fundamento da matemática quando o ensino é voltado para a formação de estatísticos, enquanto seria mais produtivo um conteúdo reduzido de matemática quando os estudantes serão, no futuro, apenas usuários dessa ferramenta. (SILVA, et al, 2002, p.219)

Isto posto, não devemos pensar que a Matemática será “banida” da Estatística uma vez que as “*ideias estatísticas são formalizadas pela Matemática*”. Hoje podemos, inclusive, contar com o auxílio de softwares estatísticos que privilegiam a análise dos dados “poupando” os alunos dos cálculos matemáticos e de seus fundamentos teóricos. É importante ressaltar que, segundo Hand (1998, apud SILVA et al, 2002, p.220), quanto maior for o conhecimento dos fundamentos matemáticos do usuário da Estatística, menor será o risco de cometer erros na análise dos dados.

Assim, acreditamos que se a Estatística for ministrada como “*disciplina de serviço*” com ênfase na análise dos dados e não nos fundamentos da Matemática, talvez possa ser mais apreciada pelos alunos. Estes, quando trazem em sua bagagem experiências desagradáveis com a Matemática, não raro transferem essa atitude negativa à Estatística e comumente optam por cursos que não tenham a disciplina na grade curricular, não a elegem como disciplina optativa ou a cursam

com muita dificuldade. Por conseguinte, sem lhe dar o devido valor, estudam-na somente para conseguir a nota mínima para aprovação.

Nesse sentido, encontramos trabalhos como o de Grácio, Oliveira e Oliveira (s.d), que apontam para o fato de que quando a Estatística - como “*disciplina de serviço*” - é ministrada a partir de uma investigação sugerida pelos alunos, envolvendo-os no processo de pesquisa e análise, ela apresenta muito mais significado para eles.

Desta forma e neste contexto, endossamos a idéia de que os alunos possam mudar a posicionamento em relação à Estatística. Se atitudes positivas ou negativas são adquiridas no decorrer dos anos, são elas também passíveis de mudanças. Silva et al (2002) argumenta que, para ocorrer mudanças positivas nas atitudes dos alunos em relação à Estatística, é necessário que o professor também esteja motivado e busque estratégias estimulantes. As autoras completam que “*no momento em que o aluno começa a perceber que está entendendo o conteúdo e está encontrando aplicação no seu cotidiano acadêmico e pessoal, é possível, então, se efetivar essa mudança de atitudes*” (SILVA, et al, 2002, p.219).

Acreditamos também que na formação profissional Superior a Estatística possa contribuir muito mais aos conhecimentos dos futuros formandos se as Instituições de Ensino e seu corpo docente acreditarem na cultura das atitudes positivas em relação à disciplina. É preciso, ainda, que haja meios de amenizar experiências desagradáveis com a Estatística ou com a Matemática, como no trabalho com a análise exploratória dos dados.

A filosofia da análise exploratória dos dados faz parte da Estatística Descritiva e é uma das três áreas da Estatística. Nessa filosofia, encontramos a leitura e interpretação de gráficos e tabelas como elementos da Estatística básica. Para esse trabalho, estamos considerando como conhecimento básico em leitura e interpretação de gráficos e tabelas as seguintes capacidades: diferenciar e saber quando usar os tipos de gráficos (barras, colunas e setor), ler informações explícitas e implícitas em gráficos e tabelas, construir gráficos e tabelas respeitando as escalas e elementos de sua construção e saber converter dados que estão representados em tabelas para gráficos e vice-versa. Voltaremos a esta questão no capítulo II, quando a discutiremos com maior profundidade.

A escolha por investigar os conhecimentos sobre leitura e interpretação de gráficos e tabelas se apóia nas idéias de GAL (2002), que afirma que tais conhecimentos são básicos em Estatística e essenciais para a formação dos que vivem em uma sociedade saturada de informações.

As pesquisas em Educação da Matemática apontam para a realidade do ensino deficiente da Estatística. Assim, a partir dos resultados de seu estudo, Vasques (2007) conclui que os alunos que entram em contato com a Estatística no curso superior encontram dificuldades na disciplina devido, provavelmente, a uma lacuna em conhecimentos estatísticos construída ainda nos tempos do Ensino Fundamental e Médio.

Já, Pereira (2007), que desenvolveu pesquisa com professores de Matemática em exercício e que já se formaram, afirma que aqueles professores pesquisados trazem deficiência na sua formação Estatística ainda dos tempos de graduação e estas deficiências refletem em suas aulas e conseqüentemente no aprendizado insuficiente de seus alunos.

Pensando em todos esses fatores é que surgiu o interesse em pesquisar como está ocorrendo o ensino da Estatística na graduação de professores, especificamente nos cursos de Licenciaturas em Matemática e em Pedagogia. Provavelmente serão eles os futuros responsáveis por difundir a Estatística estudada na graduação para os alunos do ensino básico. Concomitantemente, intencionou-se investigar o ensino da Estatística na graduação de um curso que a utiliza num contexto de usuário, como o de Administração, realizando um estudo comparativo entre os conhecimentos básicos dos alunos destes três cursos em leitura e interpretação de gráficos e tabelas.

### **1.3. OBJETIVO E QUESTÃO DE PESQUISA**

O objetivo dessa pesquisa é investigar quais são os conhecimentos básicos dos alunos dos cursos de Licenciatura em Matemática, Pedagogia e bacharelado em Administração com relação à leitura e interpretação de gráficos e tabelas estudados na disciplina de Estatística.

Baseados em Cazorla (2002), acreditamos que o ensino da Estatística nos diversos cursos acadêmicos se dá por pelo menos três motivos, a saber:

1º. Para cursos de Bacharelado em Estatística com base sólida em Estatística, a fim de capacitar o aluno a ser além de usuário da Estatística também um consultor estatístico e produtor de relatórios complexos;

2º. Para cursos de Licenciatura, com ênfase na Estatística básica focando os métodos científicos, amostragem e análise exploratória de dados, porém, num contexto de usuário ou retransmissor desta Estatística aprendida no curso acadêmico para seus futuros alunos;

3º. Para cursos acadêmicos como o de Administração, que farão da Estatística uma “*disciplina de serviço*” num contexto de usuário, com ênfase na Estatística básica e também focando os métodos científicos, amostragem e análise exploratória dos dados.

Ponderando tais motivos e a investigação de como a Estatística é ensinada no Ensino Superior para alunos que futuramente serão professores, fizemos a opção pelos cursos de Licenciatura em Matemática e Pedagogia. Por outro lado, para investigarmos o ensino da Estatística para os alunos que futuramente serão usuários da Estatística aprendida na graduação, optamos pelo curso de Administração. Nessa pesquisa não iremos trabalhar com alunos de curso de bacharelado em Estatística por não ser do objetivo de nossa pesquisa.

Dentre tantos cursos de licenciatura, a opção por Licenciatura em Matemática se deu por acreditarmos serem esses futuros professores os responsáveis por ensinar, de uma forma mais analítica, a Estatística básica nas escolas. Explica-se a escolha do curso de Pedagogia por que os atuais alunos dessa licenciatura, quando educadores, poderão ser responsáveis pela introdução da mesma Estatística, porém de forma mais pictórica e menos analítica uma vez que esta aparece como um dos eixos indicadores dos PCN, como segue:

[...] Um olhar mais atento para nossa sociedade mostra a necessidade de acrescentar a esses conteúdos aqueles que permitam ao cidadão “tratar” as informações que recebe cotidianamente, aprendendo a lidar com dados estatísticos, tabelas e gráficos, a raciocinar utilizando ideias relativas à probabilidade e à combinatória.(BRASIL, 1998, p. 49).

O bacharelado em Administração foi escolhido por entendermos ser um curso que faz da Estatística básica uma “*disciplina de serviço*”, com ênfase na leitura e interpretação de gráficos e tabelas e pela aplicabilidade em sua carreira em vários momentos.

Percebemos, então, que a Estatística não é uma área utilizada somente em cursos de bacharelado em Matemática, Licenciatura em Matemática ou Pedagogia. Sabemos que a Estatística, em alguns cursos de Educação Superior, pode torna-se uma ferramenta de trabalho muito útil e às vezes necessária, como no caso do curso de Administração. Esse fato nos leva a crer que a Estatística pode ser utilizada com a função de uma ferramenta de trabalho.

Observe-se a Figura 1 que, do nosso ponto de vista, ilustra a Estatística dentre muitos de seus diferentes usos, como uma ferramenta de trabalho levando em conta as Licenciaturas em Matemática, em Pedagogia e o bacharelado em Administração:

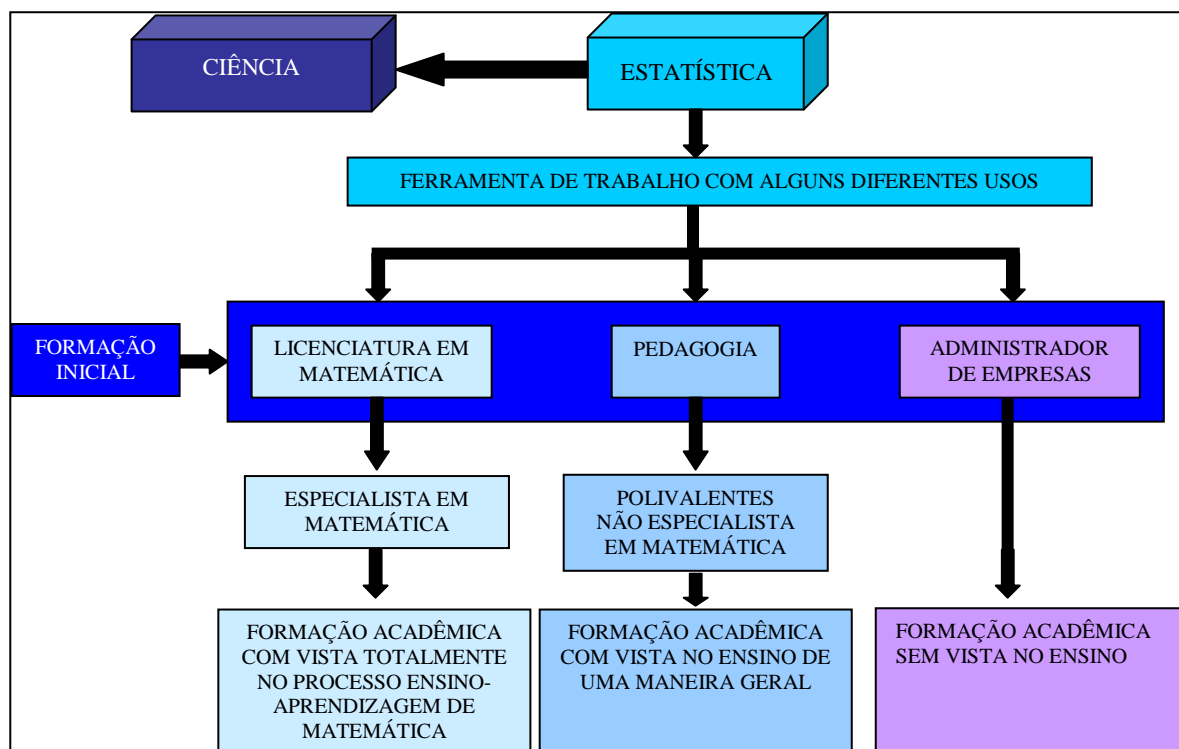


Figura 1: Esquema de alguns diferentes usos da Estatística.

Estamos considerando como “Especialistas em Matemática” os alunos do curso de Licenciatura em Matemática e futuros professores por acreditarmos que

estes apresentam fundamentos matemáticos subjacentes aos conteúdos estatísticos. Como “Polivalentes” ou “Não Especialistas” tratamos os alunos de Pedagogia que também serão professores ou administradores escolares por acreditarmos que estes utilizem os conteúdos estatísticos num contexto de usuário, sem muitos fundamentos matemáticos, apesar de serem licenciados. Tais futuros educadores farão uso da Estatística como uma ferramenta com finalidade pedagógica.

Como temos três cursos distintos e percebemos que apenas o curso de Administração não apresenta uma formação inicial voltada para o Ensino, mas com a finalidade de dar suporte ao seu trabalho, faz-se necessário a identificação de cada um dos três grupos, a saber:

- Grupo dos Futuros Licenciados em Matemática (**GFLM**)
- Grupo dos Futuros Pedagogos (**GFP**)
- Grupo dos Futuros Administradores (**GFADM**)

A escolha de fazer um estudo comparativo entre os alunos dos três cursos distintos ocorreu pelo fato de querer investigar quais são os conhecimentos básicos estatísticos de leitura e interpretação de gráficos e tabelas que foram internalizados pelos alunos desses três grupos (GFLM), (GFP) ou (GFADM), após a conclusão da disciplina de Estatística.

A luz das reflexões aqui explicitadas e sem perder de vista o objetivo de nossa pesquisa, que é o de investigar quais são os conhecimentos básicos dos alunos dos cursos de Licenciatura em Matemática, Pedagogia e Administração sobre a leitura e interpretação de gráficos e tabelas estudados na disciplina de Estatística, lançamos mão da nossa questão de pesquisa:

***QUAIS SÃO OS CONHECIMENTOS SOBRE LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS QUE ALUNOS DOS CURSOS DE LICENCIATURA EM MATEMÁTICA, PEDAGOGIA E BACHARELADO EM ADMINISTRAÇÃO MOBILIZAM AO RESOLVER SITUAÇÕES APRESENTADAS EM FORMA DE PROBLEMAS?***

Para conseguir responder a esta questão principal, precisaremos responder primeiramente algumas outras questões, de caráter mais específico, que nos permitirão ter compreensão mais abrangente dos conhecimentos básicos estatísticos dos alunos dos três grupos e, assim, obter material mais apurado para responder à questão principal. São elas:

- Qual é o nível de compreensão gráfica que o grupo dos futuros licenciados em Matemática (GFLM), o grupo dos futuros Pedagogos (GFP) e o grupo dos futuros Administradores (GFADM) apresentam?
- Qual é o nível de representação tabular, que o GFLM, o GFP e o GFADM apresentam?
- Estes três grupos de alunos, GFLM, GFP ou GFADM sabem converter dados estatísticos apresentados na forma de tabelas em gráficos e vice-versa, segundo Duval (2005)?
- Qual grupo tem melhor desempenho na conversão de dados estatísticos apresentados na forma de tabelas para gráficos e vice-versa?
- Quais dos grupos se saem melhor em questões que envolvem cálculos de medidas de tendência central: o GFLM, o GFP ou o GFADM?

Temos como hipótese e como provável resultado de nossa pesquisa que os alunos de Administração apresentarão melhor desempenho frente às situações que envolvem a leitura e interpretação de gráficos e tabelas nos testes aplicados do que os alunos de Licenciatura em Matemática e Pedagogia, mesmo tendo a ciência de que os alunos de Licenciatura em Matemática têm uma formação com mais fundamentos matemáticos subjacentes aos estatísticos do que os outros dois cursos.

Supomos que isso possa ocorrer por que possivelmente os gráficos e tabelas façam mais sentido aos alunos de Administração, uma vez que utilizam mais esses tipos de representações nas diversas disciplinas de sua jornada acadêmica.

Esses alunos apresentam, durante sua graduação, relatórios para outras disciplinas que não a de Estatística, lançando mão de gráficos e tabelas para comunicar seus resultados e provavelmente tenham uma tendência natural a gostar mais de Estatística, a ler e interpretar melhor os gráficos e tabelas do que os outros

dois cursos, que provavelmente só utilizam a Estatística nas aulas daquela disciplina.

Por esse motivo, o grupo dos alunos de Administração possivelmente traga mais base Estatística no que tange leitura e interpretação de gráficos e tabelas por se dedicar mais a tais tipos de representações de dados. Isso também pode ocorrer por apresentarem atitudes mais positivas em relação à disciplina Estatística do que os alunos de Licenciatura em Matemática e Pedagogia, que farão uso da Estatística como ferramenta pedagógica.

Tendo em mente o objetivo do estudo e preocupados em obter material teórico e prático para responder nossa questão de pesquisa, delineamos um caminho para a presente dissertação, o qual se encontra descrito resumidamente na seção a seguir.

#### **1.4. DESCRIÇÃO RESUMIDA DA DISSERTAÇÃO**

Neste primeiro capítulo apresentamos uma ponderação sobre a problemática do ensino da Estatística no ensino Superior, o objetivo do estudo e as questões de pesquisa.

No Capítulo II, encontraremos uma introdução dos conceitos básicos estatísticos como cálculo de Moda, Mediana e Média, uma discussão a cerca da Educação Estatística e de algumas idéias que permeiam esta questão, como o Letramento e o Raciocínio Estatístico, a construção e leitura de gráficos e tabelas baseado em Curcio e Wainer, as idéias de Raymond Duval e o Registro de Representação Semiótica.

O Capítulo III trará o delineamento do estudo e procedimentos aplicados, com a descrição dos sujeitos da pesquisa e do instrumento aplicado.

No Capítulo IV, discutiremos os resultados obtidos em nossos instrumentos da perspectiva de análise quantitativa e qualitativa.

E finalmente, no Capítulo V procuraremos discutir e apresentar as considerações finais baseadas em nossos estudos e no aporte teórico, procurando

responder às questões de pesquisa que serviram de impulso para essa análise. Também deixaremos algumas questões para futuras pesquisas, na tentativa de contribuir com a Educação Estatística.

### OS CONCEITOS BÁSICOS DA ESTATÍSTICA E O PERCURSO TEÓRICO DA EDUCAÇÃO ESTATÍSTICA

---

No capítulo anterior, apresentamos uma introdução sobre a importância da Educação Estatística e uma discussão sobre o cenário atual desta disciplina no ensino Superior, o objetivo de estudo e a questão de pesquisa.

Nesse capítulo pretendemos discutir os conceitos básicos da Estatística, lembrando que tomaremos como básico as medidas de tendência central (cálculo de Moda, Mediana e Média) e em seguida, discutiremos a tomada de decisão sobre quais dessas medidas utilizar. Abordaremos também os conceitos da Educação Estatística como o Letramento e Raciocínio Estatístico, a leitura de gráficos e tabelas baseados em Wainer e Curcio bem como suas construções e uma discussão acerca da Teoria de Registro de Representação Semiótica de Raymond Duval.

Na tentativa de facilitar a leitura desse capítulo, colocamos no apêndice uma discussão sobre alguns termos da Estatística que eventualmente venha a aparecer e que não é foco de nossa pesquisa, porém, necessária para a complementação da discussão dos conceitos básicos da Estatística tais como: variáveis qualitativas e quantitativas, amostra e amostragem, distribuição de frequências, regra de Sturges, diagrama de ramo e folha, fórmula de Czuber, medidas separatrizes, exemplos de cálculo da média e medidas de variabilidade.

## 2.1. MEDIDAS DE TENDÊNCIA CENTRAL: COMPREENDENDO SEUS CONCEITOS BÁSICOS

Encontramos em muitas pesquisas, notícias, informações e até mesmo em livros didáticos que não são os de Matemática, o emprego do termo “média” como meio de questionamento: Qual é a renda média anual dos metalúrgicos do ABC Paulista? Qual é a idade média dos assinantes do sexo feminino da revista “Abratamente”? Em média, quantas pessoas acreditam que existe solução para o conflito entre Palestinos e Israelenses? Quantas mulheres, em média, morrem por causa do câncer de mama em Manaus?

Então, podemos concluir que existe um número denominado “média” que representa um grupo de dados. Mas, será que existe somente um número que represente um grupo de dados? Este número é o que melhor representa-o? Existem outros números que representaria esse conjunto de dados de uma forma mais significativa?

Essas considerações nos fizeram pesquisar se existem e quais são esses números que descrevem um grupo como um todo, de maneira conveniente.

Assim, encontramos várias formas de representar um conjunto de dados, a saber, por meio de gráficos e tabelas ou pelo uso de medidas que resumem os dados: as medidas de tendência central (no qual o termo média está inserido).

Quando temos um conjunto de dados geralmente utilizamos tabelas e gráficos como meio de representá-los. Porém, podemos lançar mão de um único número que represente o que é típico, ou médio, daquele grupo. Este número, também usado nas pesquisas “...é chamado medida de tendência central, porque em geral ele está localizado mais para o meio, ou centro, de uma distribuição, onde a maior parte dos dados tende a concentrar-se.” (LEVIN; FOX, 2007, p.79).

Para Novaes e Coutinho (2008) uma forma de representar um conjunto de dados é fazendo um “retrato” desse conjunto por meio das medidas de tendência central, número esse que vêm a complementar os gráficos e tabelas. São esses números que permitem descrever de uma maneira conveniente um grupo como um todo.

Nesse sentido, temos não só o cálculo da média como sendo um número que retrata o conjunto de dados, mas também, encontramos os cálculos da moda ou mediana, a escolha de qual deles utilizar depende do objetivo da pesquisa. Então, por que o emprego do termo média é tão difundido? Será que sabemos o que este número significa?

Parece-nos que para pessoas leigas, ou mesmo para alguns alunos, o emprego do termo “média” pode em geral ser vago ou até mesmo confuso. Acreditamos que muitas pessoas empregam o termo média sem mesmo conhecer o verdadeiro significado desse número e se esta é a melhor forma de representar os dados. Encontraremos nesse capítulo, mais a frente, explicações do por que empregamos mais o termo média ao invés da moda ou mediana.

Nessa pesquisa abordaremos apenas as três medidas de tendência central mais conhecida: a Moda, a Mediana e a Média.

### 2.1.1. MODA

A **moda (Mo)** é o valor mais freqüente, o que se repete mais, o mais comum em uma distribuição de dados. Pode ser utilizada em qualquer distribuição, independentemente do nível de mensuração, entretanto, é a única de que dispomos para variáveis nominais como bacharelado em uma faculdade ou religião. Por exemplo, podemos encontrar mais protestantes nos Estados Unidos do que adeptos de outras religiões, assim, o protestantismo naquele país passa a ser a moda. Se em uma faculdade o curso mais procurado é o de bacharelado em direito, ele também representará a moda daquela faculdade.

Podemos encontrar a moda facilmente por simples inspeção: no conjunto de dados – 1, 2, 2, **5**, 1, 1, **5**, 8, **5**, 3, **5**, 3, **5**, 8, **5**, a moda é 5, por que é o valor mais freqüente (6 vezes). Devemos prestar atenção no fato de que a moda não é a freqüência do valor observado ( $f = 6$ ), mas sim o valor mais freqüente ( $Mo = 5$ )

Quando as distribuições de freqüências não têm nenhuma moda, chama-se **amodal**; quando têm duas ou mais modas são as distribuições **bimodais** e sua

representação gráfica tem dois pontos de frequência máxima, o que lembra as corcovas de um camelo, em contrapartida, com as **unimodais** (apenas um valor como moda) temos apenas um ponto de frequência máxima.

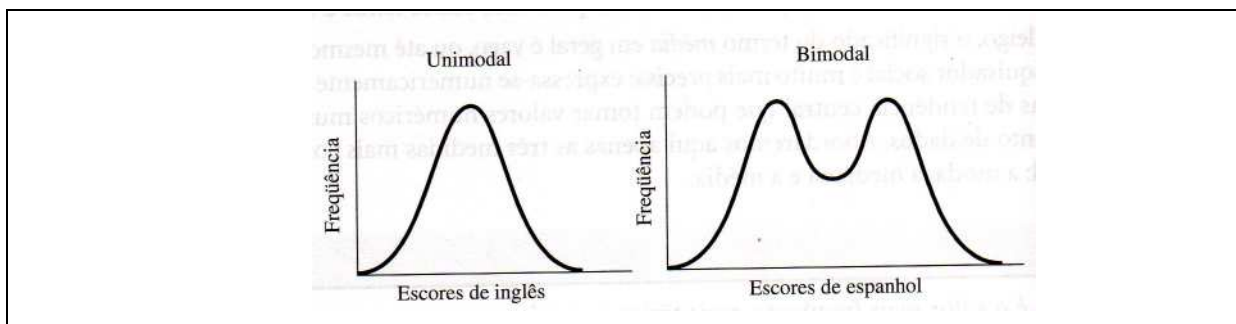


Figura 2: Representação gráfica de distribuições unimodal e bimodal de escores de testes. Retirado do livro de LEVIN e FOX, 2007, p. 80.

Outro ponto importante a ser discutido é o cálculo da moda quando os dados são agrupados em intervalos ou não. Por exemplo, considere a distribuição de 1550 peças do tipo A vendidas na primeira semana de abril de 2000 por uma loja:

Tabela 1: Distribuição de 1550 peças A, vendidas na primeira semana de abril de 2000, baseado em NOVAES e COUTINHO, 2008, p. 44.

Dias da Semana	Nº de peças	
	(f)	(%)
Segunda	215	13,9
Terça	233	15,0
Quarta	215	13,9
Quinta	244	15,7
Sexta	228	14,7
Sábado	215	13,9
Domingo	200	12,9
Total	1550	100,0

Analisando a Tabela 1, podemos observar que na segunda-feira, quarta-feira e sábado, a loja vendeu 215 peças do tipo A e como este valor se repete três vezes ( $f=3$ ) durante a semana, podemos considerar que a moda é 215 ( $Mo = 215$ ).

Agora considere a distribuição das pessoas segundo as diárias cobradas nos hotéis da cidade de “Vistajóia”, obtida por meio de uma amostra:

Tabela 2: Valor da diária escolhida pelos hóspedes da cidade Vistajóia, baseado em NOVAES e COUTINHO, 2008, p. 45.

Diária	Nº de pessoas	
	(f)	(%)
Tipo F: R\$ 50,00	37	37,0
Tipo E: R\$ 65,00	15	15,0
Tipo D: R\$ 90,00	8	8,0
Tipo B: R\$ 123,00	10	10,0
Tipo A: R\$ 215,00	30	30,0
Total	100	100,0

Observando a Tabela 2 é possível identificar que a moda do valor da diária é de R\$ 50,00 ( $M_o = R\$ 50,00$ ) porque é a frequência que mais temos pessoas. Como a porcentagem do valor mais freqüente é de 37% e sendo a amostra significativa, nas mesmas condições, se a cidade recebesse 500 hóspedes, podemos esperar que também 37% de 500 hóspedes escolheriam a hospedagem do Tipo F, isso significa que 185 pessoas iriam se hospedar neste tipo de quarto. (NOVAES e COUTINHO, 2008, p. 45).

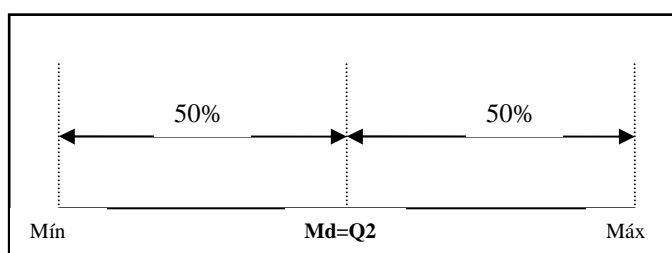
Já, a estimativa do cálculo da moda para a distribuição de frequências com dados agrupados é feita por meio de semelhança de triângulos ou a **fórmula de Czuber**, porém não iremos abordar aqui este assunto por não ser foco de nossa pesquisa. O leitor poderá encontrar a definição desta fórmula no apêndice C.

### 2.1.2. A MEDIANA

A **mediana ( $M_d$ )** é o ponto do meio de uma distribuição, podendo ser traduzida como uma medida de tendência central que separa a distribuição em duas partes iguais, em outras palavras, uma das medidas separatrizes, mais precisamente o **segundo quartil ( $Q_2$ )**. Para que seja possível encontrar este número, a primeira providência é a de “ordenar os dados, depois determinar o local no qual essa se encontra e, finalmente, calcular o valor que ela toma” (CAZORLA e SANTANA, 2006, p.21)

A posição da mediana pode ser determinada por inspeção e de uma forma intuitiva ou pela fórmula  $\frac{N+1}{2}$ , e recai exatamente no meio quando temos um número ímpar de elementos já ordenados.

Agora, quando o número de elementos é par, Levin e Fox (2007) definem que “a mediana é sempre o ponto acima do qual recaem 50% dos casos e abaixo do qual também recaem 50% dos casos”, podendo também ser classificada como o **segundo quartil (Q<sub>2</sub>)**.



Quadro 2: Demonstração da posição de Q<sub>2</sub>, adaptado de NOVAES e COUTINHO, 2008, p. 32.

Para um número par de valores, haverá dois casos médios, e a mediana estará situada exatamente a meio caminho entre os dois valores do meio.

### 2.1.3. A MÉDIA

A Estatística trabalha com alguns tipos de média: a média geométrica ou média harmônica, a média ponderada e a média aritmética. Abordaremos nessa pesquisa somente o procedimento do cálculo da média aritmética por ser o tipo de média que utilizamos em nosso instrumento de pesquisa.

Não há dúvida de que a média aritmética ( $\bar{X}$ ) é uma das medidas de tendência central mais populares e utilizadas pelas pessoas que necessitam de um número que represente um conjunto de dados. Levin e Fox (2007) definem a **média aritmética** como “a soma de um conjunto de escores dividida pelo número desses escores no conjunto”. Quando esta definição é representada como fórmula, temos:

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n} \left\{ \begin{array}{l} \text{Onde} \\ \bar{X} \text{ significa média amostral (lê-se xis barra)} \\ \sum \text{ significa somatória (letra grega maiúscula } \sigma) \\ X \text{ significa os valores brutos do conjunto de dados} \\ n \text{ significa o número total de valores do conjunto} \end{array} \right.$$

A Estatística faz uso de letras do alfabeto latino  $\bar{X}$  (média), quando trabalhamos com medidas amostrais e uso de letras minúsculas do alfabeto grego,  $\mu$  (média), quando trabalhamos com medidas populacionais, assim também temos a média como:

$$\mu = \frac{\sum_{i=1}^N x_i}{N} \left\{ \begin{array}{l} \mu \text{ significa média populacional (letra grega minúscula } \mu) \\ \sum \text{ significa somatória (letra grega maiúscula } \sigma) \\ x_i \text{ significa cada um dos valores observados} \\ N \text{ significa o número de elementos da população.} \end{array} \right.$$

#### 2.1.4. MODA, MEDIANA OU MÉDIA

A escolha da medida de tendência central a ser utilizada em uma situação particular, depende muito do nível de mensuração, da forma de distribuição dos dados e do objetivo da pesquisa. (LEVIN e FOX, 2007, p.87).

Quanto ao nível de mensuração a moda (Mo) é a medida de tendência central que pode ser aplicada tanto em variáveis qualitativas nominais, ordinais ou intervalares quanto em variáveis quantitativas discretas ou contínuas por que exige apenas uma contagem de freqüências.

A mediana (Md) exige uma ordenação crescente ou decrescente então, só podemos utilizá-la quando trabalhamos com variáveis qualitativas ordinais ou intervalares e variáveis quantitativas, mas não para dados nominais. Perceba que não teria sentido calcular a mediana da variável qualitativa filiação religiosa ou país de origem.

Já a média ( $\bar{X}$  ou  $\mu$ ) é utilizada somente para dados intervalares, nos cálculos estatísticos mais complexos. Sua aplicação em variáveis qualitativas fica sem sentido, uma vez que não calculamos a média do país de origem ou a média do gênero de vinte entrevistados.

A distribuição dos dados pode tomar duas formas: **simétrica** ou **assimétrica**. As simétricas, também chamadas de unimodal, são aquelas no formato de um “sino” e que os valores da moda, mediana e média são idênticos: o valor mais freqüente (Mo) também é o valor mais central (Md) e o centro da gravidade, o fulcro da “balança” ( $\bar{X}$ ). Nesse caso a escolha por moda, mediana ou média vai depender muito dos “objetivos específicos da pesquisa e no nível em que seus dados são medidos” (LEVIN e FOX, 2007, p. 87).

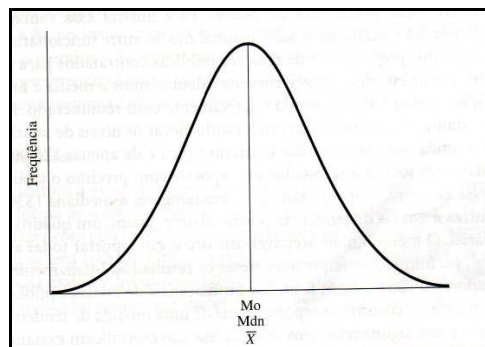


Figura 3: Exemplo de uma distribuição simétrica, unimodal, mostrando que moda, mediana e média tem valores idênticos, retidão de LEVIN e FOX, 2007, p. 87.

Já nas distribuições assimétricas os valores da moda, mediana ou média influenciam a tomada de decisão sobre qual delas utilizar. As três medidas ficam localizadas entre as “caudas” e o “pico” da distribuição sempre na ordem da moda para a mediana e para a média e não coincidem. A moda se mantém no ponto mais alto justamente por ser o valor de mais freqüência; a média se aproxima da “cauda” justamente por ser influenciada pelos valores mais extremos em qualquer uma das direções, em contrapartida a mediana sofre pequena alteração quando os extremos se modificam justamente por representar a posição mais central de uma distribuição.

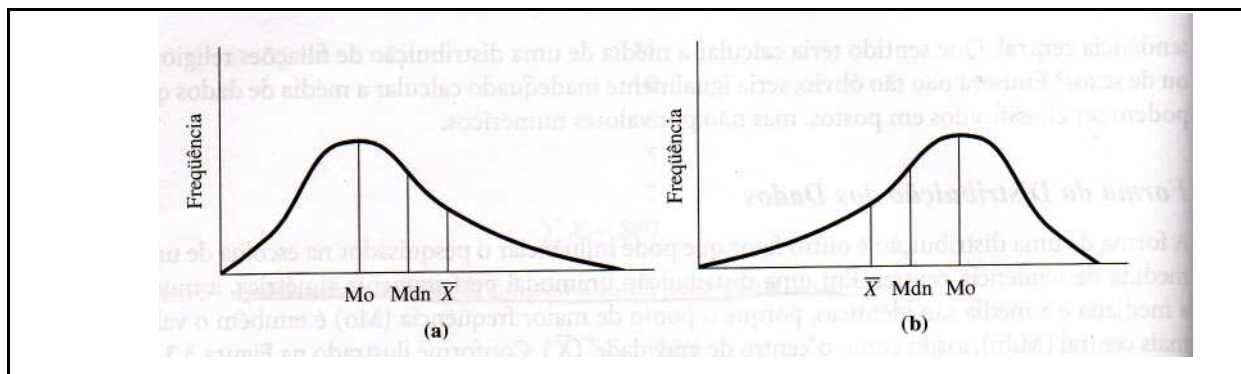


Figura 4: Exemplos das posições relativas das medidas de tendência central em uma distribuição (a) positivamente assimétrica e (b) negativamente assimétrica, retirada de LEVIN e FOX, 2007, p. 88.

Nesse tipo de distribuição a mediana se localiza sempre entre a média e a moda. Para Levin e Fox (2007) essa característica faz com que a mediana seja “a medida mais conveniente de tendência central para descrever uma distribuição assimétrica de valores”, uma vez que é a medida mais equilibrada. Agora quando a distribuição é bimodal, é aconselhável que utilize a moda por que utilizando somente a média ou mediana podemos mascarar aspectos importantes da distribuição.

Quanto ao objetivo da pesquisa, se a intenção é a de encontrar uma medida simples, rápida o ideal é que utilize a moda. Para as distribuições assimétricas, geralmente a opção é pela mediana por ser uma medida mais equilibrada. Agora, para uma distribuição aproximadamente simétrica a média é a medida que tende ser mais precisa. A razão de utilizar a média é por que este número facilita as análises estatísticas mais avançadas e por ser mais estável “no sentido de que varia menos em diferentes amostras extraídas de qualquer população” (LEVIN e FOX, 2007, p. 90).

## 2.2. O PERCURSO TEÓRICO DA EDUCAÇÃO ESTATÍSTICA

A forma como se processa a aprendizagem da Estatística e a aquisição de conhecimento é uma das grandes preocupações e interesses dos pesquisadores em Educação Matemática e Estatística.

Na busca de respostas às indagações que contemplam a Educação Estatística é que fomos buscar algumas das teorias que embasam e sustentam o ensino da Estatística nas escolas.

### **2.2.1. EDUCAÇÃO ESTATÍSTICA: UM DESAFIO**

Acreditamos que a Educação Estatística para as escolas básicas e de Ensino Superior alcançou um patamar de desafio. Nos dias atuais, cada vez mais a sociedade exige dos cidadãos a capacidade de leitura ampla da realidade vivida e de habilidades de interpretação das informações veiculadas a todo o momento, bem como o discernimento das informações certas ou erradas na hora da tomada de decisão. Então, a escola pode promover situações que desenvolvam essas habilidades nesses cidadãos por meio do ensino da Probabilidade e Estatística.

Não podemos esquecer que encontramos em qualquer empresa o emprego da Estatística desde o processo de controle de qualidade de seus produtos até a apresentação de seu faturamento. Sem contar que o ensino e aprendizagem da Estatística e Probabilidade é cada vez mais empregado em instituições de Ensino Superior nas mais diversas áreas.

Para Watts (1991, apud VENDRAMINI e BRITO, 2001), a Educação Estatística enfrenta problemas principalmente no que tange as áreas que não são estritamente matemáticas, justamente por lidar com conceitos abstratos e linguagem própria da matemática, que para muitos é ambígua e confusa, além do que ver a Matemática como linguagem que *“lida com problemas do mundo real que envolvem tomadas de decisões em condições de incerteza”* e que gera um certo medo, apreensão, insegurança e ansiedade com relação as disciplinas que necessitam da Matemática.

Então podemos supor que este sentimento negativo com relação à Matemática também gera um sentimento igual com relação à Estatística. Isto porque a Estatística costuma ser introduzida na escola por um professor de Matemática e,

portanto, facilmente associada pelo aluno como um “tipo” de Matemática. Assim, esse sentimento pode ter sua raiz na educação básica.

Essas idéias encontram respaldo em Vendramini e Brito (2001) quando argumentam que:

Em pesquisas realizadas por vários professores, Brito (1996) constatou que as afirmações dos alunos a respeito dos sentimentos negativos gerados pelas disciplinas “matemáticas” eram constantes, e que algumas dessas disciplinas eram difíceis e aversivas. Segundo a autora, este fato parece mostrar que as pessoas, de um modo geral, e os alunos de segundo grau, em particular, não gostam da Matemática e das atividades que envolvem a Matemática, sentimento que, aparentemente, se cristalizaria na universidade. (VENDRAMINI e BRITO, 2001)

Nesse sentido, acreditamos que um dos desafios da Educação Estatística consiste justamente em fazer com que os professores do ensino básico consigam fazer com que estes alunos superem tais atitudes negativas em relação à Matemática e conseqüentemente à Estatística, procurando “quebrar essa corrente” e que desperte nesses alunos a motivação de aprender a aprender.

Em experiência relatada por Vendramini (2000) com alunos ingressantes no curso de Psicologia, foi possível constatar que esses alunos tinham uma barreira inicial quanto à disciplina Estatística, ou mesmo atitudes negativas em relação a ela. Com o trabalho desenvolvido durante o período letivo, os alunos passavam a compreender melhor o significado e a importância da Estatística para o desempenho de suas futuras profissões, mas continuavam apresentando dificuldades para atingir um bom desempenho acadêmico na disciplina. (VENDRAMINI e BRITO, 2001)

As autoras ainda apontam que esse desempenho acadêmico pode ser acarretado pelas atitudes negativas em relação à disciplina gerada por experiências de aprendizagem anteriores ao Ensino Superior e que constantemente, os professores universitários apontam que provavelmente este “fracasso estatístico” possa estar relacionado com um ensino deficiente ainda nos cursos de Ensino Fundamental e Médio.

Para Vendramini e Brito (2001), a palavra atitude é empregada no sentido de apresentar

[...] uma disposição mental, dirigida a objetos, eventos ou pessoas, que assume diferente direção e intensidade de acordo com as experiências do indivíduo, e que apresenta componentes do domínio afetivo, cognitivo e motor. Essa autora considera a atitude em relação à Matemática adaptada à definição de Stagner (1937), que caracterizaria essa atitude em particular por um objeto (a Matemática), uma direção (positiva ou negativa) e uma intensidade (gostar da ou ter aversão à Matemática) [...] por um objeto (a Estatística), um direção (positiva ou negativa) e uma intensidade (gostar ou ter aversão à Estatística). (VENDRAMINI e BRITO, 2001).

Parece que este quadro de atitudes negativas em relação à Estatística já ultrapassou os muros das escolas básicas e segundo Perney e Ravid (1991, apud VENDRAMINI e BRITO, 2001) *“a Estatística é vista como um obstáculo para a obtenção de diplomas de muitos universitários, que retardam o máximo possível a matrícula nesta disciplina”*.

No entanto, as pesquisas em Educação Matemática no campo da Estatística, apontam que esse quadro pode ser revertido quando os professores são mais atenciosos no quesito Ensino da Estatística e na busca de metodologias mais eficazes no processo ensino-aprendizagem, ou seja, apontam para uma possível solução.

Isto posto, podemos dizer que outro grande desafio da Educação Estatística no século XXI é fazer com que a Estatística chegue até as escolas com menos teoria e “receitas prontas” e com mais atividades que permitam aos alunos desenvolver os conceitos tão desejados na filosofia da análise exploratória de dados, dentre eles, o letramento estatístico e o raciocínio estatístico.

### **2.2.2. O LETRAMENTO ESTATÍSTICO E O RACIOCÍNIO ESTATÍSTICO**

Uma das discussões dentro da Educação Matemática é a de estabelecer o que significa ser letrado estatisticamente. Para tanto, acreditamos que antes mesmo de definirmos o que significa letramento, seja necessário partirmos da definição de competência matemática.

Para Abrantes, Serrazina e Oliveira (1999) a competência matemática transcende a sala de aula.

[...] está relacionada com as atitudes, as capacidades e os conhecimentos relativos à matemática que, de uma forma integrada, todos devem desenvolver e ser capazes de usar, podendo identificar-se com a noção de *literacia matemática*. (ABRANTES, SERRAZINA E OLIVEIRA, 1999)

O termo *literacia* é empregado em Portugal como tradução do inglês “literacy”, já no Brasil, foi traduzida como **letramento** e cada vez mais os meios acadêmicos utilizam-se deste termo em diferentes sentidos.

Nesse sentido, utilizaremos a definição de letramento como à de Carvalho (2003), que apoiada nas idéias de Abrantes, Serrazina e Oliveira (1999) define *literacia* como:

[...] uma capacidade particular e um modo de comportamento para compreender e usar a informação nas actividades do dia-a-dia tanto em casa como no emprego ou na comunidade ao mesmo tempo que permite desenvolver os conhecimentos e potencialidades que cada pessoa possui. Porém, quando pensamos em pessoas e cidadãos competentes em Estatística, ou qualquer outra disciplina, não devemos reduzir esta competência aos seus saberes característicos devendo-se acrescentar as atitudes, os valores e as capacidades. (CARVALHO, 2003, p.43)

Porém, não estamos só pensando no letramento, mas sim no letramento estatístico. Então o que seria o letramento estatístico?

Para Cazorla (2002) não existe um consenso sobre a definição do termo letramento estatístico, muitos autores tratam desse assunto, porém quase nunca o definem. Nesta pesquisa, assumiremos a definição de **letramento estatístico** no mesmo sentido que Wallman (1993) define a alfabetização estatística:

[...] a capacidade para compreender e analisar criticamente os resultados estatísticos que permeiam o dia-a-dia de qualquer cidadão bem como a capacidade de compreender as contribuições que o pensamento estatístico têm para as decisões públicas e privadas, profissionais e pessoais. (WALLMAN, 1993, p.1)

Nesse sentido, podemos acreditar que o letramento estatístico é uma habilidade construída juntamente com o desenvolvimento do raciocínio estatístico. Mas então, o que é o raciocínio estatístico?

Para Gal e Garfield (1999, apud CARVALHO, 2003) o raciocínio estatístico é muito confundido com o raciocínio matemático, talvez pelo fato da Estatística ser ministrada pelo professor de Matemática e muitas vezes com ênfase na computação, nas fórmulas e nos procedimentos, porém, são definições distintas. Para estes mesmos autores, o **raciocínio estatístico**:

pode ser definido como sendo o modo como as pessoas raciocinam com as idéias estatísticas, conseguindo assim dar um significado à informação estatística. O que envolve fazer interpretações com base em conjuntos de dados, representações de dados ou resumos de dados. Muitos dos raciocínios estatísticos combinam dados e acaso o que leva a ter de ser capaz de fazer interpretações estatísticas e inferências. (GAL & GARFIELD, 1999, apud CARVALHO, 2003)

Então, encontramos Gal e Garfield (1997, apud CARVALHO, 2003) distinguindo o raciocínio estatístico do raciocínio matemático sob quatro aspectos:

- O raciocínio estatístico trabalha com o número num contexto e tal contexto promove o tipo de interpretação dos dados.
- No raciocínio estatístico a indeterminação dos dados distingue-se da exploração matemática mais precisa e de natureza mais finita.
- Os procedimentos da matemática fazem parte e são necessários para a construção do raciocínio estatístico, porém, não são limitados por eles.
- Os problemas estatísticos não possuem uma única solução, não conferem um *status* de completamente errados nem certos, “*devendo ser avaliados em termos da qualidade do raciocínio, da adequação dos métodos utilizados, à natureza dos dados existentes*”.

Isto posto, vale a pena ressaltar que concordamos com as idéias defendidas por Gal e Garfield (1997, apud CARVALHO, 2003) quando estes apontam os quatro aspectos em que o raciocínio estatístico distingue-se do raciocínio matemático, porém discordamos das ideias implícitas de que o raciocínio matemático não necessita de um contexto para interpretação dos números e que tenha natureza finita.

Parece-nos que os autores não levaram em consideração que o raciocínio matemático necessita sim de um contexto, mesmo que seja um contexto matemático para compreender como os números operam dentro da própria Matemática e que na educação básica talvez, o raciocínio matemático, tenha uma natureza finita, porém sabemos que a Matemática também tem uma natureza infinita.

Além disso, os autores ainda definem sete objetivos necessários para que os alunos construam o raciocínio estatístico, a saber:

- a) “*a compreensão da lógica das investigações estatísticas*”: refere-se a importância de saber sobre a organização dos dados, em outras palavras, compreender desde a escolha de população ou amostra e o tipo de inferência até a escolha dos registros gráficos a serem utilizados ou os números que representarão estes dados, passando pelo cálculo dos erros.
- b) “*a compreensão dos processos presentes numa investigação estatística*”: refere-se ao desenvolvimento de todas as etapas de uma investigação estatística de forma clara, bem definida.
- c) “*o domínio de certos procedimentos estatísticos*”: saber manipular os cálculos estatísticos como medidas de tendência central e dispersão e ter a clareza de como os resultados devem ser comunicados.
- d) “*as ligações que se podem fazer com a Matemática e quais as idéias matemática presentes nos procedimentos estatísticos*”: fazer uso da Matemática para esclarecer a alteração da média na presença de valores extremos ou à mediana quando alteramos os valores.
- e) “*a noção de probabilidade e de incerteza*”: para os autores este é um dos objetivos mais importantes, o de desenvolver nos alunos a idéia de probabilidade e incerteza por meio de atividades que criem um cenário capaz de reproduzir fenômenos imprevisíveis.
- f) “*a importância de desenvolver a capacidade de comunicar-se estatisticamente*”: criar situações que promovam a reflexão e o uso dos termos estatísticos na argumentação crítica, em outras palavras, saber fazer uma análise exploratória dos dados.
- g) “*o trabalho com a investigação estatística e o desenvolvimento de atitudes estatísticas positivas*”: a metodologia na investigação estatística na qual os alunos são envolvidos desde a escolha do tema, o levantamento da questão, a escolha dos instrumentos a utilizar, a coleta dos dados o tratamento destes e a forma de representá-los, promove uma atitude positiva nos alunos.

Nesse sentido, podemos dizer que o Raciocínio Estatístico está ligado ao Raciocínio Matemático pelo fato da Estatística usar os conceitos e procedimentos da Matemática, porém distinguem-se em alguns pontos.

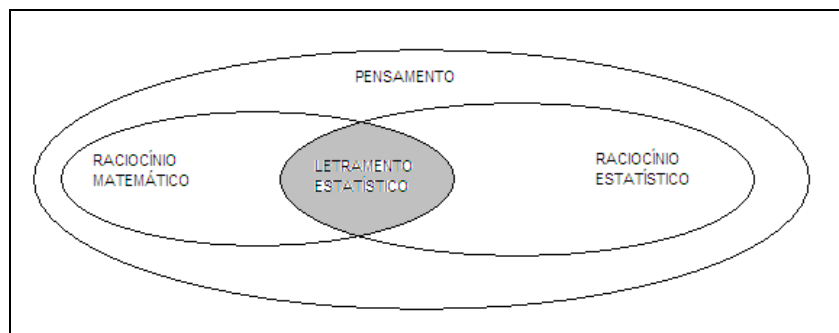


Figura 5: Esquema do funcionamento do pensamento em relação ao raciocínio estatístico e raciocínio matemático

Então, para que um indivíduo seja considerado como letrado estatisticamente, faz-se necessário que este desenvolva seus raciocínio estatístico e conseqüentemente que apresente alguns conceitos estatísticos como o de ler e construir gráficos e tabelas, entre outros.

### **2.2.3. A CONSTRUÇÃO, LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS**

A leitura e a interpretação de gráficos e tabelas não passa despercebido pelas pesquisas em Educação Matemática. Contamos com várias pesquisas pautadas nesse tema, citamos como exemplo os trabalhos de Cazorla (2002), Araújo (2007), Ribeiro (2007) e Silva (2008) que voltaram o foco de suas pesquisas para o campo da leitura e/ou interpretação de gráficos e/ou tabelas.

Nesse sentido, nossa pesquisa visa contribuir com esse campo de estudo, procurando para isso, embasamento teórico que nos dê argumentos para a análise do material de estudo.

Consideramos ser importante, antes mesmo de propormos um estudo sobre a literatura disponível sobre gráficos e tabelas, apresentar a definição do termo tabela e gráfico na Estatística:

- **Tabela:** “é um quadro que resume um conjunto de observações” (GUEDES, et al., s.d.)
- **Gráfico:** “é uma representação simbólica de dados, geralmente relacionando duas ou mais variáveis, utilizando o sistema de coordenadas cartesianas” (LEINHARDT; ZASLAVSKY; STEIN, 1990, apud CAZORLA, 2002).
- **Gráfico de Setor:** é um diagrama<sup>2</sup> representado na forma circular onde os valores de cada categoria levantada são proporcionais às respectivas medidas dos ângulos.

Segundo Cazorla (2002) é essencial compreendermos que o gráfico estatístico difere do gráfico de funções matemáticas, a saber:

Enquanto as funções matemáticas e seus respectivos gráficos modelam funções determinísticas, do tipo  $Y = F(X)$ , ou seja, dado um valor para  $X$  pode-se conhecer exatamente o valor de  $Y$ . Ao contrário, os gráficos estatísticos modelam funções não determinísticas, do tipo  $Y = F(X) + \varepsilon$ , onde  $\varepsilon$  representa o erro aleatório e é formado pelo componente aleatório, devido ao processo de amostragem; pelo erro explicado pela ausência de variáveis que podem estar interferindo no comportamento da primeira e pelos erros de medida dos instrumentos. Existem, também, outros gráficos estatísticos que não necessariamente estão preocupados com a modelagem da relação entre as variáveis, mas apenas pretendem ilustrar o comportamento das mesmas. Assim, pode-se classificar os gráficos em dois grandes tipos: os *matemáticos* e os *estatísticos*. (CAZORLA, 2002, p.45)

Assim, para essa pesquisa iremos abordar apenas os gráficos do tipo estatístico.

A seguir apresentaremos algumas considerações sobre as regras para a construção de tabelas e gráficos bem como os trabalhos de autores que versam sobre leitura e interpretação destes.

---

<sup>2</sup> Diagrama: Representação gráfica de determinado fenômeno.

### 2.2.3.1. CONSTRUÇÃO DE TABELAS

Como em nosso material de estudo só focamos as tabelas do tipo simples e de dupla entrada, aqui iremos explicar somente esses dois tipos de tabelas.

A tabela chamada simples apresenta as categorias das variáveis observadas em um estudo e suas respectivas contagens, em outros termos, o número de ocorrências de uma categoria anotadas em uma coluna denominada de frequência absoluta.

São classificadas em temporal, geográfica, específica ou categórica e comparativa. Uma tabela simples é temporal quando as observações levam em consideração o tempo, é geográfica quando as variáveis fazem alusão ao local da ocorrência; é específica quando apresentam o local e tempo fixos, e, comparativa quando a tabela apresenta informações de duas ou mais variáveis.

A tabela comparativa é um caso especial de tabela simples e mais conhecida como tabela de dupla entrada ou mais, ou ainda tabela cruzada.

Toda tabela deve conter alguns elementos que simplificam, organizam e proporcionam uma consulta rápida aos dados. São eles:

- Título
- Cabeçalho
- Classificação (Temporal, Geográfica, Específica, Comparativa)
- Número de linhas e colunas
- Fonte (quando retirada de algum trabalho)
- Análise

Tabela rs – Título respondendo as perguntas: o quê, onde e quando?	
Coluna indicadora	Cabeçalho
Conteúdo da linha	Célula
	Coluna
Fonte: Origem dos dados.	
Nota: Informação esclarecedora.	Corpo da tabela

Quadro 3: Generalização da distribuição de uma tabela simples retirada de Guedes et al, s.d., p.6

Observando a Quadro 3, percebemos que essa contém os elementos necessários para a construção de tabela simples: apresenta um título - respostas de meninos à retirada de um brinquedo, um cabeçalho com a variável – resposta da criança e a frequência absoluta  $f$ , duas colunas bem definidas e seis linhas. Em trabalhos científicos, não utilizamos as bordas verticais das colunas e nem as linhas onde anotamos as categorias das variáveis e sua frequência absoluta, porém conservamos as linhas horizontais que separam o título do cabeçalho, o cabeçalho do corpo da tabela, o corpo do total e o total da fonte.

<b>TABELA 2.1 Respostas de meninos à retirada de um brinquedo</b>	
<b>Resposta da criança</b>	<b><math>f</math></b>
Choro	25
Raiva	15
Retirada	5
Procura por outro brinquedo	5
	$N = 50$

Figura 6: Exemplo da estrutura de uma tabela simples, retirada de Levin e Fox, 2007, p. 28.

Já na tabela da Figura 7, encontramos os dados dispostos em uma tabela de dupla entrada. Perceba que esse tipo de tabela necessita de mais colunas uma vez que a intenção é promover uma comparação entre as distribuições de frequências de duas variáveis.

TABELA 2.2 Respostas de crianças à retirada de um brinquedo, por sexo		
Resposta da criança	Sexo da criança	
	Masculino	Feminino
Choro	25	28
Raiva	15	3
Retirada	5	4
Procura por outro brinquedo	5	15
Total	50	50

Figura 7: Exemplo da estrutura de uma tabela de dupla entrada, retirada de Levin e Fox, 2007, p.29.

Para Duval (2003) as tabelas significam uma forma simples de representar as informações e são largamente utilizadas não só em livros ou artigos, mas também na publicidade, como uma das principais formas de comunicação escrita.

Para esse autor, existe uma grande diversidade de tabelas e essa diversidade é refletida em três aspectos. O primeiro diz respeito sobre a “*disposição formal dos seus ‘micro-espaço’*”, em outros termos, a construção da estrutura de cada tabela. Podemos encontrar tabelas sem margens e figurantes de um cenário simples ou tabelas mais complexas e portadoras de um “*emaranhado de linhas e colunas tão complicadas*” e ainda aquelas tabelas que podem ser acondicionadas em caixas. O segundo aspecto sobre a diversidade é que aos alunos, desde o Ensino Fundamental até o Ensino Superior, são oferecidos os três tipos de tabelas sem a preocupação de explorar ao máximo suas disposições devido à “*crença na simplicidade de comunicação*” que pode ser destruída até mesmo com um simples questionamento ao professor. O terceiro aspecto diz respeito à formação dos adultos, que sabem da importância de dominar a leitura/interpretação e construção de tabelas, mas que “*grande parte desse público é quase ‘analfabeto’ nessas práticas*” (DUVAL, 2003, p.3-6)

Nesse sentido, acreditamos que as concepções sobre a construção de tabelas é também um ponto importante a ser trabalhado, uma vez que sua leitura e interpretação também dependem dessa habilidade.

Para Wainer (1992) uma apresentação tabular melhora significativamente quando seguimos três passos que orientam para a construção de uma tabela que tem o intuito de comunicar e não somente armazenar os dados. São elas:

- a) **“Ordenar fileiras e colunas de uma maneira que faça sentido”** – estruture os valores da tabela em ordem decrescente e quando temporal, sempre do passado para o futuro.
- b) **“Arredondar os valores”** - os humanos não entendem facilmente e nem memorizam mais que dois algarismos e além de que, estatisticamente dois algarismos já são suficientes para representar um número.
- c) **“Linhas e colunas são importantes”** – o espaçamento entre as colunas e entre as linhas favorecem a percepção do fato que pretendemos demonstrar.

### 2.2.3.2. LEITURA E INTERPRETAÇÃO DE TABELAS

A leitura e interpretação de tabelas não é uma tarefa tão imediata por mais simples que esta se apresente, evoca uma função cognitiva e desenvoltura visual. Para Duval (2002, apud FLORES e MORETTI, 2006, p.31) a contribuição cognitiva das tabelas e suas diferentes utilizações só são válidas quando analisada sob dois pontos: a organização representacional das tabelas, ou seja, como as informações são dispostas dentro das colunas e linhas e a função cognitiva que ela desempenha.

A disposição das linhas e colunas de uma tabela distribui os dados e por meio de cruzamento das mesmas, facilita a visualização e a identificação da informação, porém para Duval (2002, apud FLORES e MORETTI, 2006, p.32) o funcionamento representacional de uma tabela é muito mais que isso, é necessário diferenciar as *“especificidades das tabelas em relação às outras representações gráficas”* (FLORES e MORETTI, 2006, p.31).

Nesse sentido, Howard Wainer (1992) descreve três passos que definem uma apresentação tabular de qualidade (item 3.1.2.1 dessa pesquisa) e ainda estabelece uma estrutura teórica para gráficos, fruto de uma revisão de Bertin (1973) e que pode ser generalizada e empregada *“na medida de numerações com apresentações em forma de tabela”* (WAINER, 1992, p.18).

Para Wainer é essencial que primeiramente tenhamos ciência do tipo de perguntas gráficas ou tabulares que utilizaremos a fim de medir o nível de leitura que um indivíduo se encontra, a saber:

a) **“Nível básico”** – é o nível em que as questões somente extraem da tabela os dados que estão explícitos;

b) **“Nível intermediário”** – é o nível em que as questões exigem a interpolação ou a percepção da relação existente entre os dados de uma tabela.

a) **“Nível avançado”** – é o nível em que as questões abordam um maior entendimento das estruturas dos dados em sua totalidade, comparando tendências, analisando questões implícitas e privilegiando a visão global da tabela.

Para Wainer (1992), do ponto de vista da construção é comum encontrar tabelas “pobres” que contemplam no máximo cinco questões e que geralmente exploram somente o nível básico. O autor ainda argumenta que o nível de dificuldade exigido em questões criadas a partir de uma tabela, refere-se geralmente à manipulação algébrica dos valores constantes nesta e que apenas solicitam *“passos múltiplos do mesmo nível”* ao invés de aprofundar o nível de inferência.

### **2.2.3.3. CONSTRUÇÃO DE GRÁFICOS**

Os gráficos estão entre os recursos visuais mais utilizados nos meios de comunicação e embora, às vezes, forneçam menos dados que uma tabela, os gráficos favorecem a visão global dos dados, evidenciando as tendências, as ocorrências ocasionais, os valores máximos e mínimos e as ordens de grandezas dos fenômenos.

A escolha pelo tipo de gráfico que iremos utilizar depende da natureza dos dados. Quando as variáveis são qualitativas ou quantitativas de dados discretos, podemos lançar mão dos gráficos de barras e colunas para promover uma análise comparativa entre as partes ou lançar mão do gráfico de setor quando desejamos ter uma visão “parte-todo”. Agora, quando a variável é quantitativa contínua

representada em uma distribuição com os dados agrupados, podemos utilizar o gráfico chamado histograma e por meio deste traçar o polígono de freqüências.

Para os gráficos de colunas e barras, devemos trabalhar com os dois eixos de um gráfico cartesiano: construir o eixo das abscissas (horizontal) e o das ordenadas (vertical). A proporção do eixo vertical é aproximadamente de 60% a 70% da largura do eixo horizontal. No final de cada eixo anotamos as “*unidades utilizadas nas escalas que mensuram as grandezas representadas*”, em geral no eixo vertical anotamos a freqüência relativa e no eixo horizontal, as categorias das variáveis. Também apresenta um título ou legenda explicativa de maneira breve e clara e sua elaboração segue uma regra básica, ser capaz de responder a três exigências: o quê, onde e quando. Preferencialmente, não abusar na variação das cores de um mesmo gráfico, salvo no caso de barras e colunas comparativas que utilizamos cores diferentes para cada variável. (GUEDES, et al, sd, p. 18)

O exemplo<sup>3</sup> a seguir mostra a construção de um gráfico de colunas e suas particularidades.

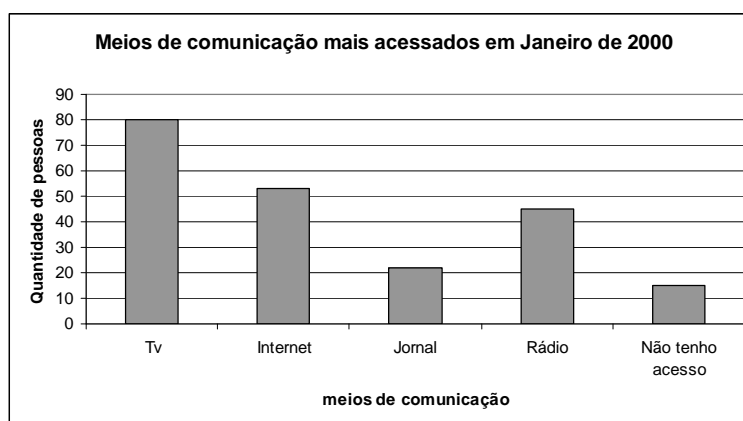


Figura 8: Exemplo da estrutura da construção de um gráfico de colunas com uma pesquisa fictícia realizada com 215 pessoas da cidade Vistajoia sobre qual o meio de comunicação mais utilizada no período de janeiro de 2000.

Podemos observar na Figura 8, a presença do título (Meios de comunicação mais acessados em janeiro de 2000) que responde a regra básica da elaboração de um título, o eixo vertical que indica a quantidade de pessoas e o horizontal onde encontramos as categorias da variável e a construção do retângulo correspondente

---

<sup>3</sup> Exemplo fornecido por nós.

à frequência observada. Note que se trata de uma única variável, então, cada retângulo correspondente à categoria estabelecida deve manter a mesma cor.

Já a Figura 9<sup>4</sup> a seguir foi construída com base nos mesmos dados acima e segue os mesmos preceitos.



Figura 9: Exemplo da estrutura da construção de um gráfico de barras com uma pesquisa fictícia realizada com 215 pessoas da cidade Vistajoia sobre qual o meio de comunicação mais utilizada no período de janeiro de 2000.

Também podemos lançar mão do gráfico de colunas ou de barras quando pretendemos promover uma comparação entre as variáveis, porém no lugar da frequência absoluta é recomendada a utilização da frequência relativa (porcentagem), uma vez que permite a comparação entre as variáveis, a despeito da ordem de grandeza.

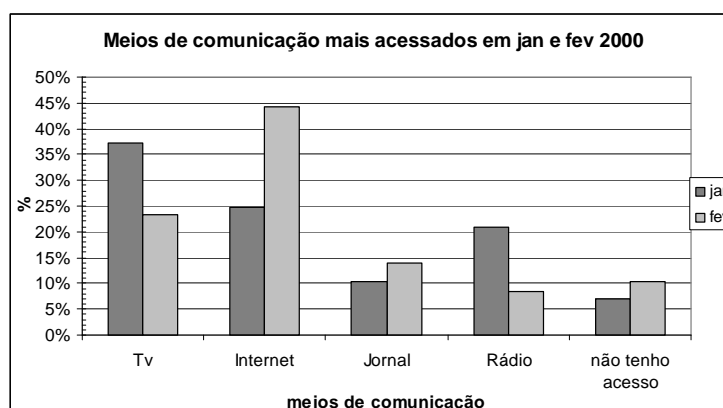


Figura 10: Exemplo de gráfico de colunas promovendo a comparação de uma pesquisa realizada com 215 pessoas da Cidade Vistajoia sobre qual o meio de comunicação mais utilizada no período de jan e fev de 2000.

<sup>4</sup> Exemplo fornecido por nós.

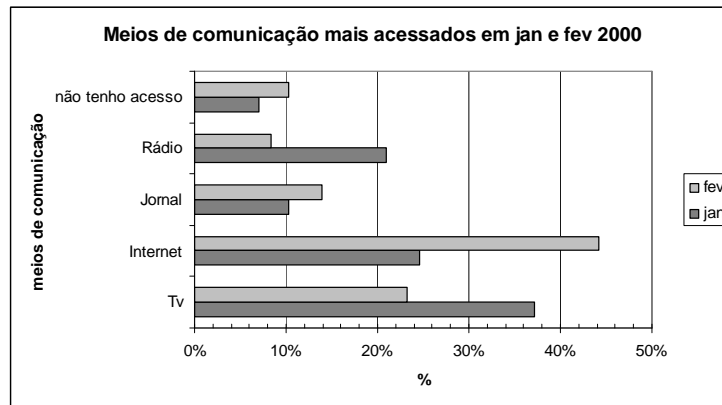


Figura 11: Exemplo de gráfico de barras promovendo a comparação de uma pesquisa realizada com 215 pessoas da Cidade Vistajoia sobre qual o meio de comunicação mais utilizada no período de jan e fev de 2000.:

Agora, para a construção de um gráfico de setor, projetamos um círculo de raio qualquer e dividimos em tantas partes quanto forem as categorias e com áreas proporcionais às frequências de cada categoria. O uso deste tipo de gráfico é recomendado quando não temos um número muito grande de categorias e estas não apresentam uma ordem definida. Para o cálculo das áreas proporcionais às frequências utilizamos uma regra de proporção simples entre a respectiva frequência e o valor total (em graus) da circunferência:  $\frac{n^{\circ} total}{f} = \frac{360^{\circ}}{x}$

O exemplo<sup>5</sup> da Figura 12, mostra a construção de um gráfico de setores e suas particularidades.

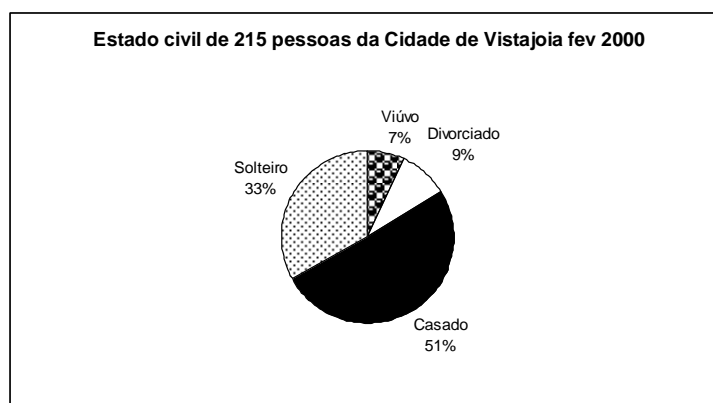


Figura 12: Exemplo da estrutura de um gráfico de setor de uma pesquisa fictícia realizada em fev 2000 sobre o estado civil de 215 habitantes da Cidade de Vistajoia.

<sup>5</sup> Exemplo fornecido por nós

Ainda para as variáveis qualitativas podemos encontrar a representação dos dados também como gráficos de linha. Já para as variáveis quantitativas discretas temos os gráficos de frequências acumuladas e bastões. Para as variáveis quantitativas contínuas podemos optar pelo histograma, polígono de frequência, diagrama de ramo-e-folhas ou gráfico de frequência acumulada ou ogiva. Porém, nessa pesquisa não iremos tratar da construção desses gráficos uma vez que nosso foco é o trabalho com gráficos de setor, coluna e barras.

#### 2.2.3.4. LEITURA E INTERPRETAÇÃO DE GRÁFICOS

Os gráficos representam um suporte de comunicação de dados que já foram coletados, organizados e analisados. Tais gráficos necessitam de uma leitura capaz de interpretar as informações neles contidos.

Para Flores e Moretti (2005) “*a leitura exige por parte do leitor certa intimidade, e também domínio, do modo de representação utilizado*”, ou seja, para fazer uma boa leitura é necessário saber construir. Por isso não é suficiente que os leitores dessas informações apenas leiam, mas também construam gráficos e tabelas. Estas autoras acreditam que as representações gráficas desencadeiam as quatro funções cognitivas do pensamento designadas por Duval (1999), a saber:

- **Função de comunicação:** função de transmissão e recepção de mensagens e códigos estabelecidos entre os indivíduos.
- **Função de tratamento:** transforma um tipo de representação em outro tipo.
- **Função de objetivação:** é a função de tomada de consciência do “não sabido” sobre o assunto abordado e a exteriorização das informações apreendidas e por muitas vezes confundida com a função de comunicação.
- **Função de identificação:** é a função de localização, de encontro das informações no meio de tantas outras e que diz respeito à organização das informações na memória.

Então, quando pensamos nas funções cognitivas do pensamento desencadeadas em relação aos gráficos temos a função de **identificação** que é a primeira a ser mobilizada pelo leitor de uma representação gráfica, uma vez que é preciso localizar no gráfico as informações pedidas dentre muitas outras. Temos ainda a função de **comunicação** dos gráficos que salta aos nossos olhos em qualquer jornal impresso ou falado, livros, revistas, etc e as funções de **tratamento e objetivação** da representação gráfica que podem ser trabalhadas no cotidiano escolar dentro do ensino da Estatística juntamente com a comunicação e identificação.

Já com relação à sistematização de leitura e interpretação de gráficos, encontramos o trabalho de Jacques Bertin (1967, apud CAZORLA, 2002, p. 54) como autor de “*uma taxonomia dos componentes gráficos e das propriedades do sistema perceptual, introduzindo uma gramática para a descrição de gráficos*” que foi revisada por Curcio no ano de 1987.

Curcio (1989) em sua revisão definiu três níveis de compreensão gráfica, a saber:

- **Leitura dos dados**
- **Leitura entre os dados**
- **Leitura além dos dados.**

Para esse autor o nível de compreensão “**leitura dos dados**” é aquele em que o indivíduo é capaz de compreender somente os fatos explícitos que observa nos dados, não existe interpretação e o nível cognitivo exigido para realizar tal tarefa é muito baixo.

Para ilustrar o que significa “leitura dos dados” vamos exemplificar com um gráfico de setores e questões que abordam esse nível de compreensão:

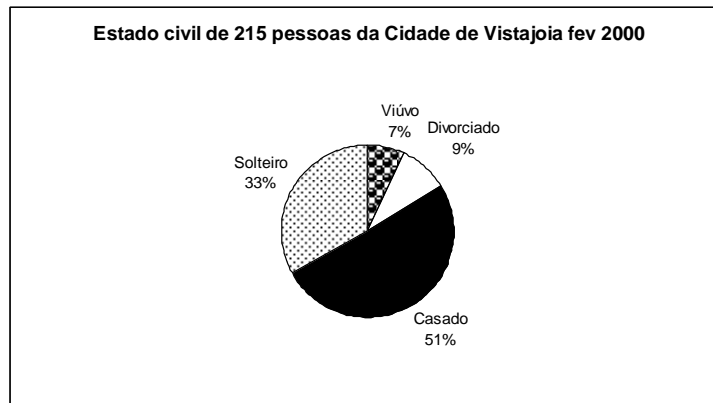


Figura 13: Estado civil de 215 pessoas da Cidade de Vistajoia em fev 2000

- Quantas pessoas foram pesquisadas?
- Qual a porcentagem de pessoas casadas? E de viúvos?
- Qual o mês e ano da pesquisa?
- Qual a porcentagem de pessoas solteiras?

Já o indivíduo se enquadra no nível de compreensão “**leitura entre os dados**” quando é capaz de interpretar o dados e fazer uso de outras habilidades matemáticas como adicionar, subtrair, multiplicar e dividir os dados, a fim de comparar quantidades, combinar e integrar as informações contidas no gráfico. Este é o nível de compreensão mais fácil de ser encontrado nas diversas atividades escolares e nos diferentes tipos de testes.

Observe o gráfico da Figura 14 e as questões que exigem o nível de compreensão “leitura entre os dados”:

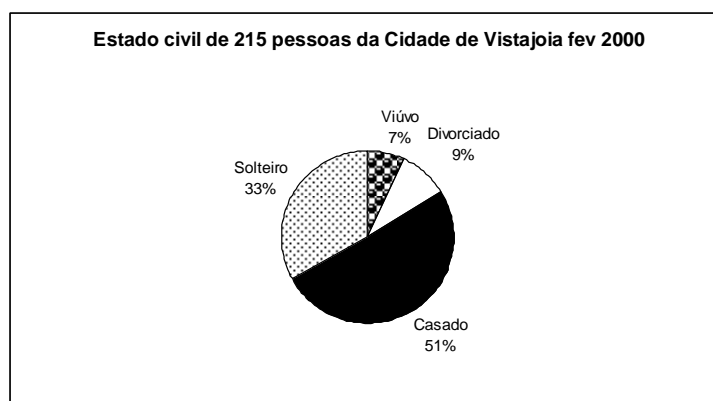


Figura 14: Gráfico de setor com estado civil de 215 pessoas da Cidade de Vistajoia em fev 2000.

- a) Qual a diferença entre o número de pessoas casadas e solteiras?
- b) Qual o total de pessoas solteiras, viúvas e divorciadas da pesquisa?

O nível de compreensão “**leitura além dos dados**” é o nível em que o leitor não encontra explícito no gráfico dados suficientes para a solução do seu problema, então busca os conhecimentos adquiridos existentes em sua memória. Também é o nível que demanda certa autonomia do leitor, uma vez que este faz inferências sobre os dados, porém baseados em informações que não estão no gráfico, mas sim, resgatando informações de aprendizagens anteriores. Observe o gráfico da Figura 15 que trabalha com o nível “leitura além dos dados”:

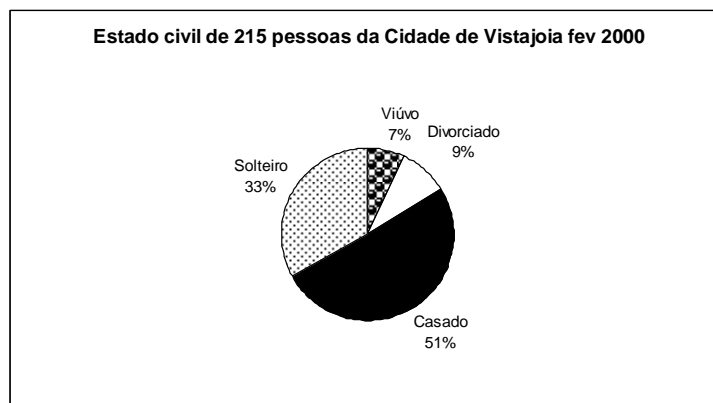


Figura 15: Estado civil de 215 pessoas da Cidade de Vistajoia em fev 2000

- a) Vamos considerar que 215 pessoas seja uma amostra representativa da população da Cidade de Vistajoia. Nas mesmas condições, se a pesquisa fosse realizada com 1350 pessoas, qual o percentual de solteiros e casados que podemos esperar?
- b) Este percentual representa quantas pessoas solteiras e casadas?

#### 2.2.4. REGISTROS DE REPRESENTAÇÃO SEMIÓTICA: A TEORIA DE RAYMOND DUVAL

Filósofo e psicólogo francês, Raymond Duval desenvolveu fundamentos da Psicologia Cognitiva e que culminou em sua obra *Sémiosis et pensée humaine*.

A obra de Duval (2005) é essencialmente uma análise do conhecimento matemático, do sistema de produção das representações semióticas deste conhecimento. As representações semióticas são “*produções constituídas pelo emprego de signos pertencentes a um sistema de representação os quais têm suas dificuldades próprias de significado e de funcionamento*”. (DUVAL, 1993, apud DAMM, 1999, p.143).

O raciocínio matemático é viabilizado por meio das representações semióticas e toda produção e comunicação em Matemática se dá neste sistema. Apesar da teoria de Duval não estender-se à Estatística podemos dizer que o raciocínio estatístico também comunga das representações semióticas.

Para Damm (1999), a Matemática se comunica por meio de objetos abstratos e que não são diretamente acessíveis à percepção como “*conceitos, propriedades, estruturas e relações que podem expressar diferentes situações*”, fato este que nessa pesquisa estendemos também à Estatística. Então, se faz necessário lançar mão de diversas formas de representar um mesmo objeto matemático e ou estatístico por meio de símbolos, signos, códigos, tabelas, gráficos, algoritmos e desenhos que permite a comunicação entre a atividade cognitiva e os sujeitos.

É comum considerar que as representações semióticas registradas na manipulação da Matemática seja a exteriorização das representações mentais, porém, Duval (1993, apud DAMM, 1999) classifica esta como uma visão ingênua e superficial, pois as representações semióticas são também necessárias para as atividades cognitivas do pensamento, em outras palavras, sem as representações semióticas o sujeito não constrói o conhecimento matemático aprendido. Ele chama de “*semiósis*” a apreensão ou a produção de uma representação semiótica e de “*noésis*” a apreensão conceitual de um objeto, que segundo este autor, um ainda não pode ser separado do outro.

Essas considerações podem também ser exemplificadas<sup>6</sup> com a Estatística, foco de nossa pesquisa: considere o conjunto de dados com a variável qualitativa “gosto pela Matemática” dos alunos do 1º ano A do Ensino Médio da escola “Ratatui” e seus diferentes tipos de representação:

a) Representação por meio de Tabela:

Tabela 3: Distribuição de frequência do gosto pela Matemática de 50 alunos do 1º ano do Ensino Médio da escola Ratatui.

Gosto pela Matemática	Nº de Alunos	%
Não	8	16,0
Pouco	12	24,0
Regular	24	48,0
Muito	6	12,0
Total	50	100,0

b) Representação por meio de Gráfico de setor:

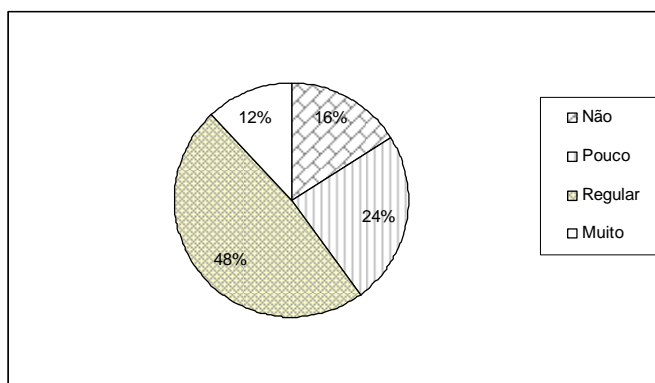


Figura 16: Gráfico de setor sobre o gosto pela Matemática de 50 alunos do 1º ano do Ensino Médio da escola Ratatui.

c) Representação por meio de Gráfico de barras:

<sup>6</sup> Exemplo fornecido por nós.

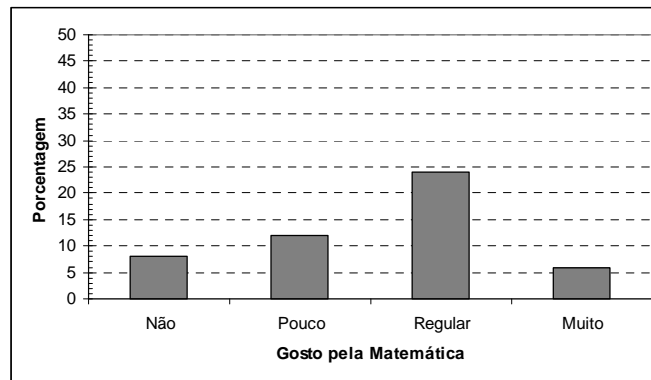


Figura 17: Gráfico de colunas sobre o gosto pela Matemática de 50 alunos do 1º ano do Ensino Médio da escola Ratatui.

Podemos observar que temos dois tipos de representação - representação tabular (Tabela 3) e representação gráfica (Figura 16 e 17) - do mesmo objeto matemático (variável qualitativa – gosto pela Matemática). Baseados na teoria de Duval observamos que o fato do aluno saber representar o conjunto de dados por meio de uma tabela ou gráfico (*semiósis*) não garante que ele tenha a apreensão do conceito (*noésis*) do objeto matemático ‘variável qualitativa’.

Portanto, a apreensão conceitual dos objetos matemáticos (*noésis*) somente será possível com a coordenação de vários tipos de representação (*semiósis*). Quanto maior o repertório de representações semióticas de um mesmo objeto matemático ou estatístico, maior a contribuição para o funcionamento cognitivo do sujeito aprendiz.

Para Duval (1999, apud, Almouloud, 2007) um *registro* de representação é um sistema semiótico de função cognitiva e que se diferencia dos códigos. Os *registros* trazem em seu bojo toda a representação direta de um conteúdo do conhecimento, por exemplo, quando vemos o registro  $f(x)$  imediatamente associamos este registro com o conteúdo “função” e todas suas implicações matemática.

Para que um sistema semiótico seja considerado um registro de representação é necessário que este contemple as três atividades cognitivas fundamentais ligadas à “*semiósis*” a saber: a **formação de uma representação identificável**, o **tratamento** e a **conversão**.

**1) A formação de uma representação identificável:** é estabelecida por meio de um enunciado compreensível que permita a identificação do objeto matemático.

Nesse sentido, esta grande variedade de representações semióticas foi organizada em quatro grupos como no Quadro 4:

	<b>Representação Discursiva</b>	<b>Representação Não-Discursiva</b>
<b>Registros Multifuncionais:</b> Os tratamentos não são algoritmizáveis.	Língua Natural Associações verbais (conceituais). Forma de raciocinar: <ul style="list-style-type: none"> <li>• Argumentação a partir de observações, de crenças...,</li> <li>• Dedução válida a partir de definição ou de teoremas.</li> </ul>	Figuras geométricas planas ou em perspectivas (configurações em dimensão 0, 1, 2 ou 3). <ul style="list-style-type: none"> <li>• Apreensão operatória e não somente perceptiva;</li> <li>• Construção com instrumentos.</li> </ul>
<b>Registros Monofuncionais:</b> Os tratamentos são principalmente algoritmos.	Sistemas de Escritas: <ul style="list-style-type: none"> <li>• Numéricas (binária, decimal, fracionária, etc.); <ul style="list-style-type: none"> <li>• Algébricas;</li> </ul> </li> <li>• Simbólicas (línguas formais); <ul style="list-style-type: none"> <li>• Cálculo;</li> </ul> </li> </ul>	Gráficos cartesianos. <ul style="list-style-type: none"> <li>• Mudanças de sistemas de coordenadas;</li> <li>• Interpolação, extrapolação.</li> </ul>

Quadro 4: Classificação dos diferentes registros mobilizáveis no funcionamento matemático (fazer matemático, atividade matemática) retirado de Duval (2005, p. 14).

Como exemplo de tipos diferentes de representação semiótica em Estatística, podemos citar a representação de um conjunto de dados na forma de tabela, de gráfico de setor, de barras ou como medidas de tendência central.

**2) O tratamento:** é a transformação da representação em outra representação, porém ficando no mesmo registro. Além disso, o tratamento está ligado à forma interna de como iremos manipular o registro e não com o objeto matemático propriamente dito.

Exemplos de tratamentos:

a) Registro algébrico para registro algébrico

$$5x + 4 = 3x - 2$$

$$5x - 3x = -2 - 4$$

$$2x = -6$$

$$x = -3$$

b) Registro numérico fracionário para registro numérico fracionário (nesse caso o tratamento é a simplificação)

$$\frac{25}{100} = \frac{1}{4}$$

c) Pensando em Estatística podemos considerar como tratamento a passagem de Registro tabular de um rol de dados estatísticos (tabela 6) para Registro tabular de uma distribuição de freqüências (Tabela 4). Perceba que continuamos no mesmo tipo de registro “tabelas”.

- Um levantamento foi feito com 10 alunos da escola “Ratatuí” sobre suas disciplinas favoritas e representadas em uma tabela.

Tabela 4: Disciplinas favoritas de 10 alunos da escola Ratatuí

<b>Alunos</b>	<b>Disciplinas favoritas</b>
Cicília	Língua Portuguesa
Adriana	Língua Portuguesa
Dulcimara	Geografia
Renata	Língua Portuguesa
Edmar	Matemática
Claudete	História
Cássia	Educação Física
Carmen	Educação Física
Tamura	Geografia
Cleide	Ciências

Baseados na teoria de Duval, realizaremos o tratamento da tabela acima em uma tabela de distribuição de frequências, ficando no mesmo registro “tabelas”.

Tabela 5: Distribuição das disciplinas favoritas de 10 alunos

<b>Disciplinas favoritas</b>	<b>Nº de alunos</b>
Língua Portuguesa	3
Geografia	2
Matemática	1
História	1
Educação Física	2
Ciências	1
<b>TOTAL</b>	<b>10</b>

Ainda podemos citar como tratamento dentro da Estatística, o fato de mudarmos a escala de um gráfico, causando um aumento ou diminuição na

amplitude das classes consideradas. Estaríamos dentro do mesmo registro, o registro gráfico, porém com um tratamento.

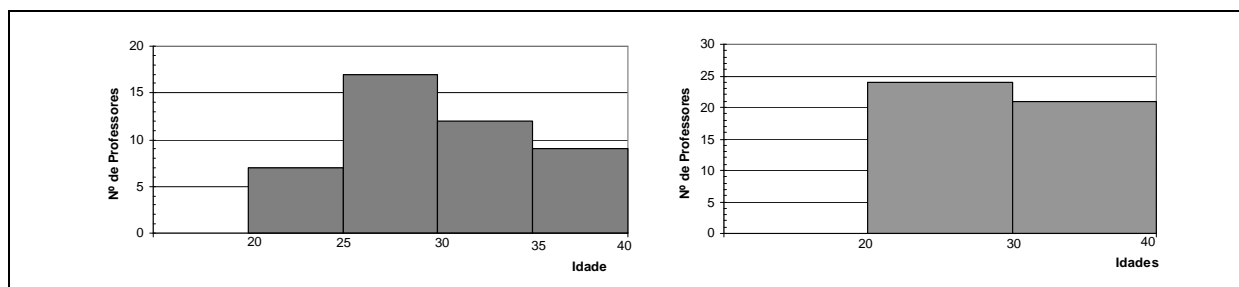


Figura 18: Mudança da escala do gráfico da variável idade

**3) A conversão:** é a transformação de um registro de representação em outro registro conservando a totalidade ou parte do objeto matemático trabalhado.

Exemplos de conversão:

a) Registro numérico decimal para registro numérico fracionário

$$0,5 = \frac{1}{2}$$

b) Registro numérico decimal para registro numérico exponencial

$$0,005 = 5 \cdot 10^{-3}$$

c) Em Estatística podemos considerar como conversão a representação dos dados de uma tabela de distribuição de frequência da disciplinas favoritas de 10 alunos da escola “Ratatuí” (tabela 6) em representação gráfica (figura 19).

Tabela 6: Distribuição das disciplinas favoritas de 10 alunos

<b>Disciplinas favoritas</b>	<b>Nº de alunos</b>	<b>%</b>
Língua Portuguesa	3	30,0
Geografia	2	20,0
Matemática	1	10,0
História	1	10,0
Educação Física	2	20,0
Ciências	1	10,0
<b>TOTAL</b>	<b>10</b>	<b>100,0</b>

Agora, apresentaremos os dados acima com a conversão para registro gráfico de setor.

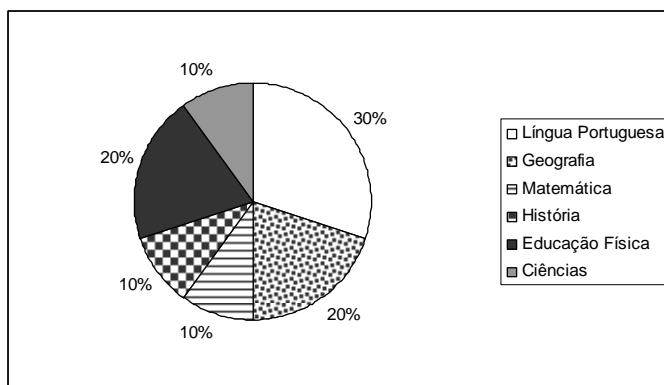


Figura 19: Gráfico de setor da distribuição das disciplinas favoritas de 10 alunos.

Para Damm (1999) o problema no ensino da Matemática, no que tange ao Registro de Representação Semiótica, está justamente no fato de que os alunos ficam trabalhando somente nas várias representações de um objeto matemático e o que garante a “noésis” é justamente a coordenação entre estes vários registros de representação, uma vez que “*é a articulação dos registros que constitui uma condição de acesso à compreensão em matemática, e não o inverso, qual seja, o ‘enclausuramento’ de cada registro*” (DUVAL, 2005, p. 22).

Isto posto, podemos inferir que um aprendiz terá apreendido realmente o objeto matemático quando este for capaz de coordenar pelo menos dois tipos de registro de representação e efetuar a mudança de registro.

Analisando a natureza cognitiva da atividade do tipo “conversão” observamos dois fenômenos: **a congruência e a não-congruência.**

- 1) **As variações de congruência e de não-congruência:** A congruência aparece quando a representação de saída sugere ou “transparece” qual deve ser a conversão efetuada para que o aluno chegue à representação final, em outras palavras, a conversão aproxima-se de uma simples codificação. Já, a não-congruência é exatamente o contrário, ou seja, o registro de saída não “transparece” qual deverá ser a conversão efetuada para que se chegue ao registro final.

- 2) **A heterogeneidade dos dois sentidos de conversão:** refere-se à importância do sentido da conversão. Quando é solicitado que o sujeito realize uma conversão do registro de saída para um outro registro final, este a realiza sem muitas complicações, porém, quando solicitado que realize a conversão do registro final para o inicial, percebe-se que estes já não reconhecem mais a situação apresentada.

Vamos tomar como exemplo de 'heterogeneidade dos dois sentidos de conversão' a conversão do registro algébrico da função  $f(x) = x + 3$  para o registro gráfico. Parece-nos que é mais simples a realização da conversão nesse sentido em que o exemplo é dado, ou seja, do registro algébrico para o gráfico, uma vez que basta escolher valores e substituir na equação para obter os pares ordenados.

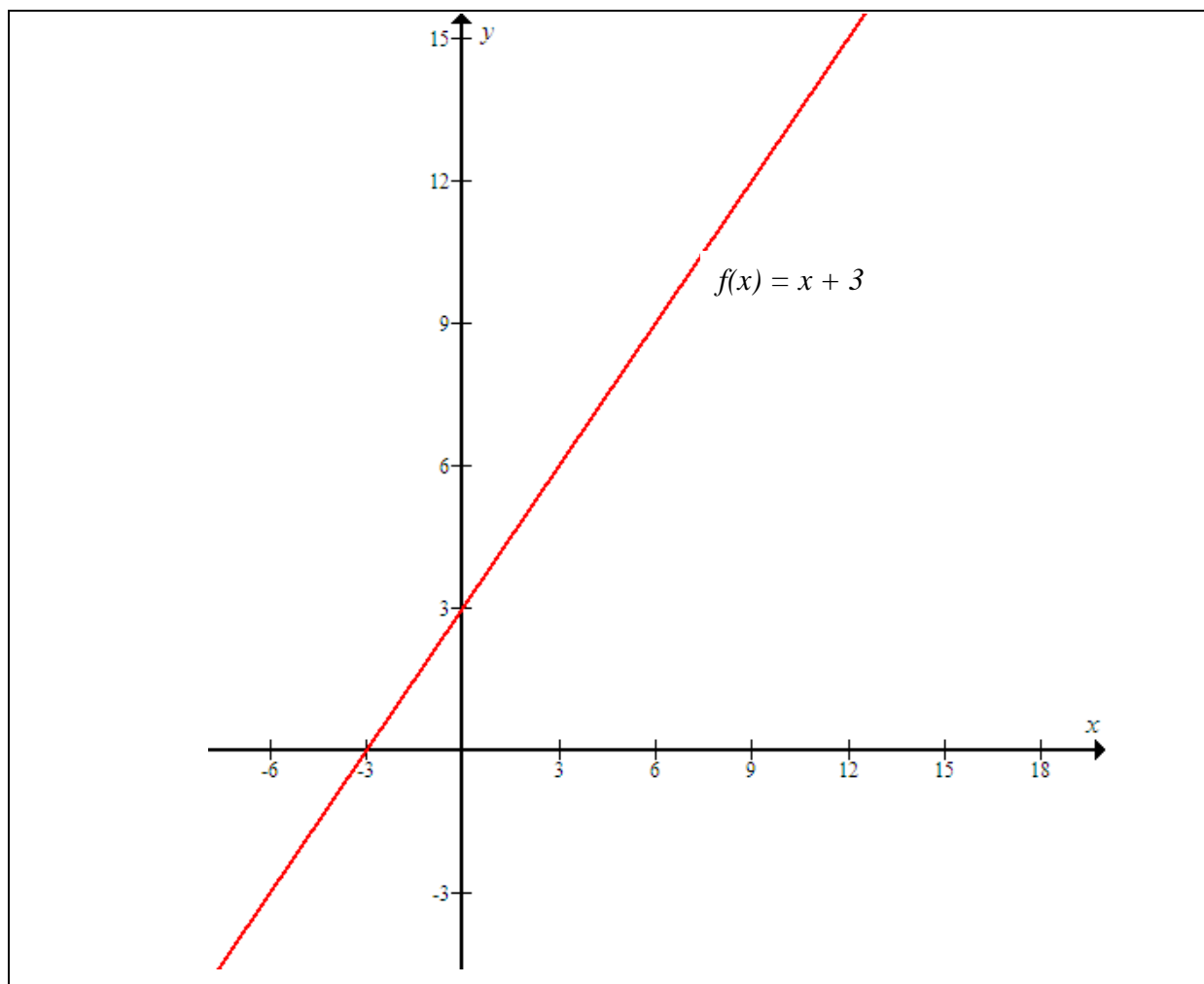


Figura 20: Gráfico cartesiano da função  $f(x) = x + 3$

Já, quando é solicitada a conversão no sentido 'inverso', do registro gráfico para o registro algébrico, podemos ter um número pequeno de acertos.

Quando é solicitada uma conversão do registro algébrico para o registro gráfico, podemos dizer que houve uma congruência, por ser uma conversão realizada de forma direta pelos alunos, porém, quando solicitamos a conversão do registro gráfico para o registro algébrico, podemos dizer que houve uma não-congruência, uma vez que esta conversão não é feita de forma direta pelos alunos.

Acreditamos que estes dois fenômenos descritos por Duval em sua teoria também pode ser 'transportada' para a Estatística.

Em Estatística a conversão de um registro de tabelas simples para um registro gráfico de barras ou colunas ou vice-versa, pode ser classificada como congruente, uma vez que é uma passagem simples, de leitura direta.

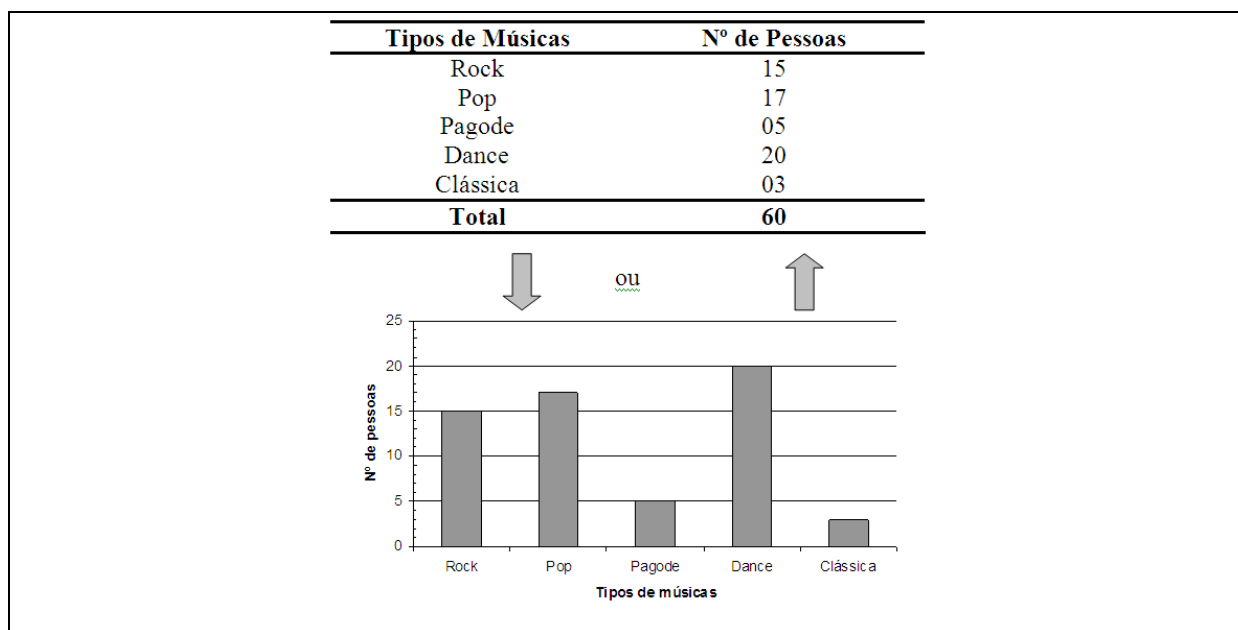


Figura 21: Demonstração de uma conversão congruente entre registro de tabela simples para gráfico de barras

Porém, no caso de uma conversão de um registro gráfico de setor para o registro de tabela de distribuição de frequências, sabemos que esta, pode ser classificada como não-congruente. Essa conversão exige que o leitor faça o cálculo da porcentagem pelo ângulo da variável apresentada ou no caso de conversão do registro tabular para o registro gráfico, é necessário que transforme a frequência em ângulos. Esta não é uma leitura direta, portanto não-congruente.

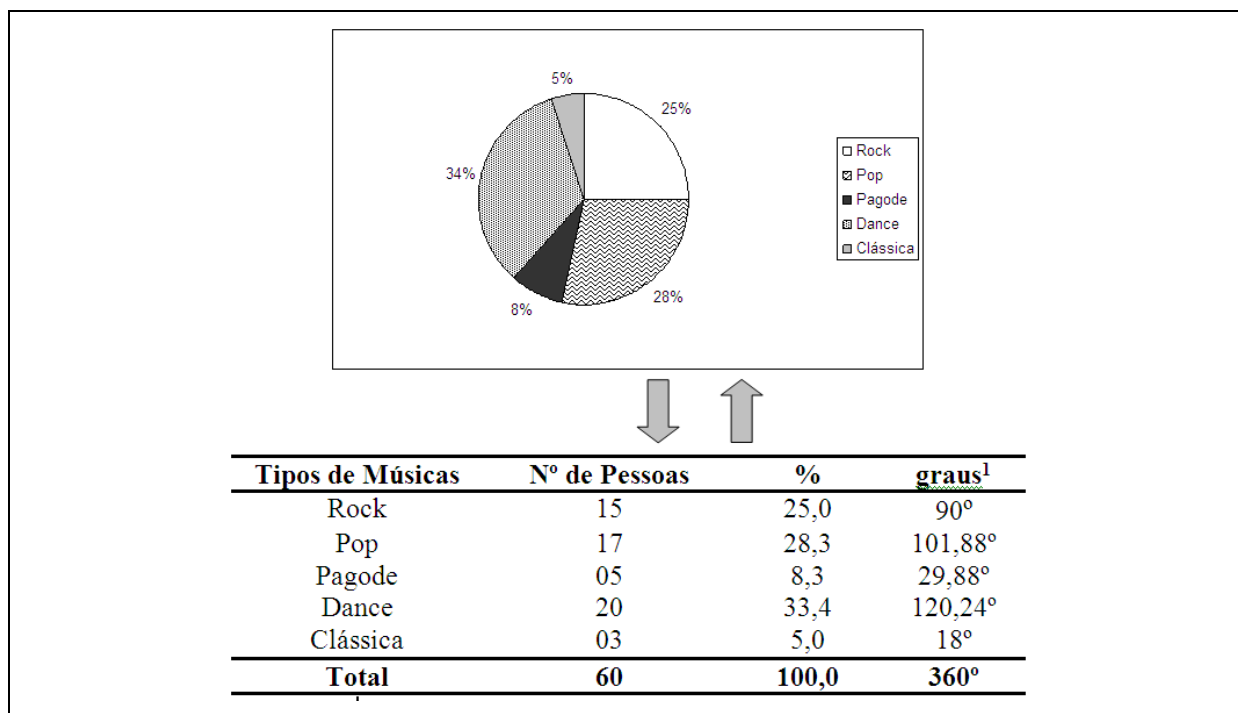


Figura 22: Demonstração de uma conversão não-congruente entre registro de gráfico de setores para tabela

Para Duval (2005) as variações de congruências e as não-equivalências dos sentidos das conversões não são resultados da compreensão do conceito do objeto matemático, pois se assim fosse, os alunos não encontrariam dificuldades com as variações de não-congruência ou com aquelas do sentido da conversão. Ele ainda afirma que “constata-se, além disso, que a não-congruência pode levar os alunos a verdadeiros bloqueios que eles não superam verdadeiramente” (DUVAL, 2005, p. 22).

Esse fato nos leva a refletir sobre o quão estão sendo trabalhados os dois sentidos da conversão e as variações de não congruência em Estatística e se a forma como estão apresentando os dados para os alunos é suficiente para que eles realmente tenham o domínio e transitem entre os diferentes tipos de registros estatísticos.

O PASSO A PASSO DA METODOLOGIA DE ESTUDO

---

No capítulo antecedente procuramos traçar alguns conceitos que consideramos básicos da Estatística e da Educação Estatística como o letramento e raciocínio estatístico, a construção, leitura e interpretação de gráficos e tabelas e a teoria de Raymond Duval, dando início a uma discussão sobre esses temas.

Já, nesse capítulo, procuraremos discutir sobre a teoria metodológica utilizada como aporte para a nossa pesquisa, apresentar o universo de estudo, os sujeitos, o material de estudo e sua aplicação.

### 3.1. A TEORIA METODOLÓGICA: PASSO A PASSO

Para compreendermos a teoria-metodológica que fomenta nosso estudo, acreditamos ser necessário, primeiramente, abordar o significado de pesquisa científica. Optamos, em meio a tantas, pela definição de Fiorentini & Lorenzato (2006) quando argumenta que:

[...] a pesquisa é um processo de estudo que consiste na busca disciplinada/metódica de saberes ou compreensões acerca de um fenômeno, problema ou questão da realidade ou presente na literatura o qual inquieta/instiga o pesquisador perante o que se sabe ou diz a respeito. (FIORENTINI & LORENZATO, 2006, p. 60).

Nessa busca disciplinada de saberes e com vista no aspecto teórico-epistemológico, nossa pesquisa rumou para uma **abordagem empírico-analítica**. A abordagem empírico-analítica desenvolve um 'papel simbólico' igual ao de uma *“máquina fotográfica: capta um determinado instante de um fenômeno sem poder*

*abarcando a evolução histórica do mesmo (...) limitando-se à evidência dos fatos num determinado momento” (FIORENTINI & LORENZATO, 2006, p.68).*

Nesse sentido, consideramos ser esse o argumento que sustenta a opção por essa abordagem, uma vez que o objetivo desse trabalho é investigar quais são os conhecimentos básicos de um grupo de alunos dos cursos de Licenciatura em Matemática, Pedagogia e Administração sobre a leitura e interpretação de gráficos e tabelas estudados na disciplina de Estatística.

Partindo desse argumento e tendo como objetivo principal a investigação de saberes estatísticos básicos, acreditamos que nossa pesquisa adquire uma perspectiva **descritiva**, uma vez que pretendemos buscar aspectos relevantes sobre o objeto de estudo e promover uma comparação entre os diferentes saberes dos três grupos de estudantes do Ensino Superior. Essa escolha se apóia nas idéias de Fiorentini e Lorenzato (2006) quando define que:

Uma pesquisa é considerada *descritiva* quando o pesquisador deseja descrever ou caracterizar com detalhes uma situação, um fenômeno ou um problema. Geralmente esse tipo de investigação utiliza a observação sistemática (não etnográfica) ou a aplicação de questionários padronizados, a partir de categorias previamente definidas. (FIORENTINI & LORENZATO, 2006, p. 70).

Isto posto, partiremos para a fase da observação e aplicação dos testes, efetuando assim o processo de coleta de dados. Quanto ao tipo de coleta, optamos pela pesquisa naturalista ou de campo onde o objeto de estudo é colhido diretamente em seu ambiente e “(...) *observados, sem intervenção e manuseio por parte do pesquisador*” (SEVERINO, 2008, p. 123).

Outro ponto a ser analisado é a constituição dos respondentes. Tomemos como exemplo o pesquisador social que de uma forma geral, procura pesquisar, analisar e tirar conclusões sobre questões que envolvem grandes números de indivíduos. Dessa forma, um pesquisador investiga uma população ou universo, o que em muitos casos, se torna inviável quer seja por tempo ou recursos limitados.

Assim, nossa pesquisa irá trabalhar com uma amostra não-aleatória, classificada como amostragem intencional, aquela que se “*baseia exclusivamente no que é conveniente para o pesquisador*” (LEVIN e FOX, 2007, p. 178).

Quanto ao tamanho da amostra, Moscarola (1990) apud Almeida et al, s.d., ressalta que “*com uma amostra inferior a 30 observações se tem chances de*

*encontrar tanto um valor errôneo ou defasado como um valor se aproximando da realidade*”, porém quando aumentamos as observações para 100 ou mais, aumentam também as chances de chegar a “*valores ou resultados alinhados com a realidade*”, do ponto de vista dos cálculos estatísticos. Definimos assim, inicialmente um número de no mínimo 150 (cento e cinquenta) respondentes.

Quanto à parte prática da coleta de dados, Marconi e Lakatos (2008) classifica a técnica como “*um conjunto de preceitos ou processos*” utilizados na obtenção dos propósitos de uma pesquisa. A parte prática da coleta de dados, ou seja, a técnica é dividida em dois grupos: “*documentação indireta*” e “*documentação direta*”. Essa última, documentação direta, se subdivide em: “*observação direta intensiva*” e “*observação direta extensiva*”. Para nossa pesquisa usaremos a observação direta extensiva que apresenta, entre outras técnicas, a do tipo **teste**. Essa escolha foi baseada nos objetivos de nossa pesquisa e na definição das mesmas autoras: “*testes são instrumentos utilizados com a finalidade de obter dados que permitam medir o rendimento, a freqüência, a capacidade ou a conduta de indivíduos de um grupo*”. (MARCONI e LAKATOS, 2008, p. 111).

Partindo desse argumento, acreditamos que tal teste poderá ser responsável pela articulação entre a proposta metodológica escolhida e a perspectiva descritiva, uma vez que este desenvolve o papel de condutor do conhecimento a ser pesquisado. É no momento da análise das respostas do teste que o pesquisador dialoga com os fenômenos estudados procurando promover a integração dos pressupostos teóricos e metodológicos com as informações produzidas, permitindo assim, um exame intensivo dos dados, suas características, amplitude e profundidade.

Então, para que possamos compreender melhor os resultados obtidos e levantar dados para responder a nossa questão de pesquisa principal e secundária, iremos adotar para a análise dos testes aplicados uma **abordagem qualitativa** quando analisarmos as estratégias utilizadas pelos respondentes e uma **abordagem quantitativa** quando levantarmos os percentuais de acertos e erros em cada questão, levando em consideração os teóricos escolhidos para dar suporte às unidades de análise de nossa pesquisa.

A luz de todas essas considerações e também buscando orientação pelas nossas questões e pelo que queremos investigar, passamos para a construção do delineamento do universo de estudos, seus sujeitos, o material de estudo e de sua aplicação.

### **3.2. O DELINEAMENTO DO UNIVERSO DE ESTUDO**

O universo de estudo foi constituído sempre com vistas na coleta de material para uma pesquisa naturalista ou de campo onde a quantidade de respondentes foi obtida por meio de uma amostra não aleatória e as instituições de ensino superior, bem como os alunos, escolhidos pela acessibilidade a eles.

No que tange a abordagem da análise das respostas corretas ou não corretas dos testes de nossa pesquisa, seguiremos o modelo de estudo quantitativo descritivo, uma vez que pretendemos descrever os fenômenos numericamente, no que tange aos percentuais de acertos e erros.

Quanto à análise das estratégias de resolução utilizadas pelos alunos e interpretação dos dados, pretendemos realizar uma abordagem qualitativa, criando possíveis categorias de classificação que possibilite descrever os fenômenos estudados por meio de palavras.

Para tanto, vamos definir quem são os sujeitos respondentes e o material de pesquisa.

#### **3.2.1. OS SUJEITOS**

O grupo de 174 (cento e setenta e quatro) sujeitos dessa pesquisa são estudantes de graduação, advindos dos cursos de Licenciatura em Matemática, Pedagogia e Bacharelado em Administração do ano letivo de 2008 e que durante seu percurso acadêmico, já cursaram a disciplina de Estatística.

A disciplina de Estatística oferecida pela instituição do curso de Administração visa de um modo geral, introduzir ao aluno a formação técnico-científica para a atuação na administração das Instituições, usando a estatística como ferramenta estratégica para a tomada de decisões. Os conteúdos programáticos são: séries estatísticas; gráficos estatísticos; variável discreta e contínua; distribuição de frequência; medidas de tendência central para dados agrupados, média aritmética, moda bruta e de Czuber, entre outros. Para os outros cursos não tivemos acesso a ementa da disciplina de Estatística.

Contamos com a participação de 57 pessoas do gênero masculino, 116 feminino e 1 que não respondeu o questionário de perfil; desses, cerca de 63% estão entre a idade de 18 e 30 anos.

Quando perguntados sobre quais os conteúdos estatísticos estudados na graduação, cerca 53% não lembram quais foram, e, dos 47% que lembram os conteúdos, os mais votados foram Moda, Média e Mediana seguido de Probabilidade e Gráficos, sendo esses dois últimos apontados como conteúdos mais importantes estudados por cerca de 30% dos sujeitos pesquisados.

Dentre os sujeitos pesquisados, encontramos 36% que lecionam e desses, 49 sujeitos ministram aulas de Matemática no Ensino Fundamental ou Ensino Médio. Quanto ao tempo de magistério, 21% estão entre 6 a 10 anos e 35% há mais de 10 anos.

Quando perguntados sobre a importância do ensino de Estatística, 93% dos pesquisados o consideram importante ou muito importante e apontam que a razão dessa importância é que a Estatística auxilia na leitura e interpretação de dados veiculados nos meios de comunicação em seu dia a dia (41%) e para tomada de decisão voltada para a área do trabalho (30%).

A pesquisa conta com 174 (cento e setenta e quatro) sujeitos distribuídos da seguinte forma: 72 (setenta e dois) sujeitos do curso de Licenciatura em Matemática, 48 (quarenta e oito) sujeitos do curso de Pedagogia e 54 (cinquenta e quatro) sujeitos de Administração.

Dentro dos 72 (setenta e dois) sujeitos do curso de Licenciatura em Matemática temos 18 (dezoito) sujeitos de uma Instituição de Ensino Superior do município de São Bernardo do Campo, Estado de São Paulo - que nessa pesquisa

chamaremos de instituição  $\alpha$  - e 54 (cinquenta e quatro) sujeitos de uma Instituição de Ensino Superior do Estado da Bahia - que chamaremos de instituição  $\beta$ .

Para a turma de Pedagogia, temos 48 sujeitos respondentes de uma Instituição de Ensino Superior do município de Registro no Estado de São Paulo.

Já para as turmas de Administração temos um total de 54 (cinquenta e quatro) sujeitos respondentes de uma Instituição de Ensino Superior do município de São Bernardo do Campo, Estado de São Paulo distribuídos da seguinte forma: 21 (vinte e um) sujeitos do 6º semestre turma C e 33 (trinta e três) sujeitos do 6º semestre turma D.

Sendo assim, observe a Figura 23 que representa o desenho dos sujeitos da pesquisa:

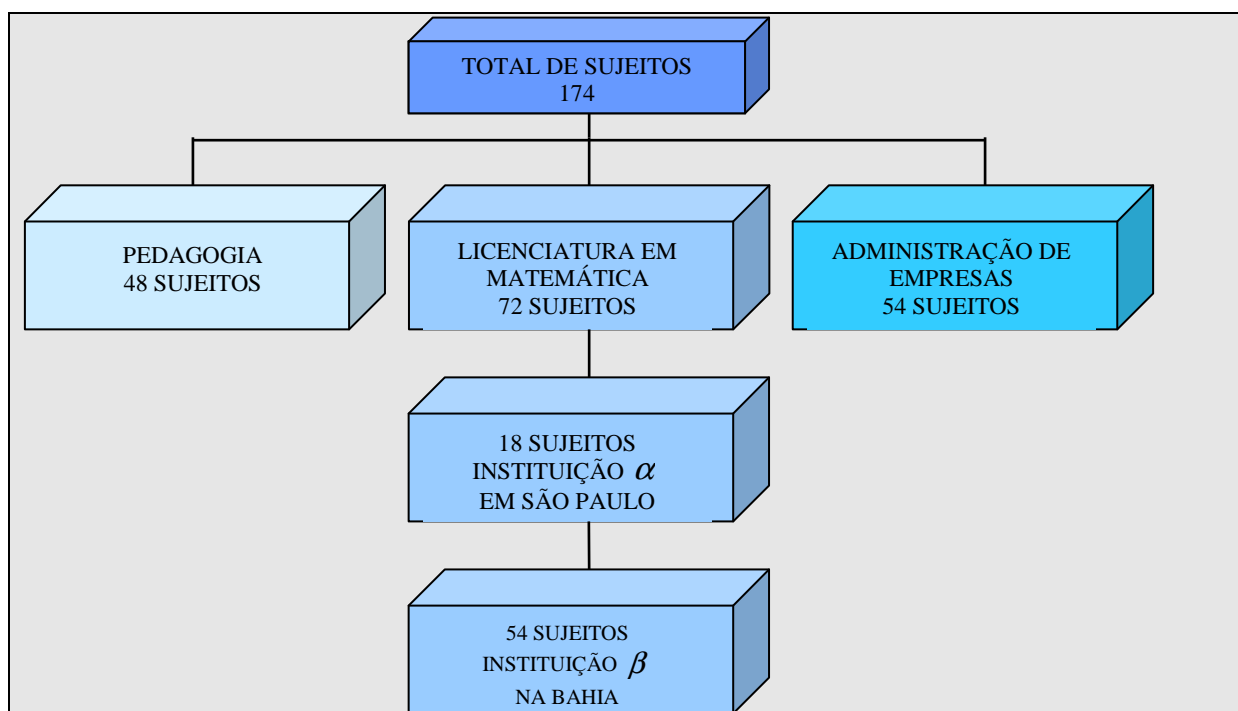


Figura 23: Desenho dos sujeitos da pesquisa

Todos os sujeitos responderam ao questionário e ao teste em suas instituições de ensino - salvo 01 (um) sujeito da instituição na Bahia que entregou o questionário sobre o perfil totalmente em branco.

Nesse sentido, teremos então os 174 (cento e setenta e quatros) sujeitos do teste divididos em três grupos distintos, a saber:

- Grupo dos Futuros Licenciados em Matemática (**GFLM**);
- Grupo dos Futuros Pedagogos (**GFP**);
- Grupo dos Futuros Administradores (**GFADM**).

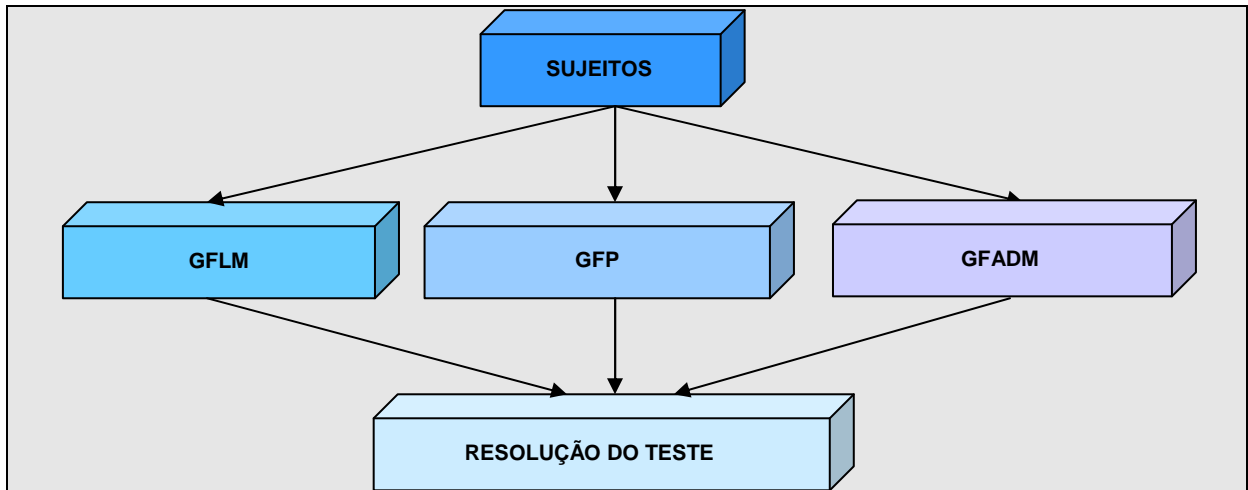


Figura 24: Desenho da pesquisa

A seguir, encontraremos uma explanação do material de estudo aplicado nos sujeitos respondentes.

### 3.2.2. O MATERIAL DE ESTUDO

O material de estudo foi composto por um caderno de seis páginas como mostra o Apêndice A em folhas do tipo A4, dividido em duas partes, a saber: na primeira parte (folha 1) encontramos um questionário com uma série de perguntas que definem o perfil de cada um dos sujeitos (questão 1 composta pelos itens 1.1 até 1.16) e da segunda parte (folha 2 até folha 6) um teste com questões que abordam conceitos básicos de Estatística (questão 2 até questão 10).

Sendo assim, o teste foi composto por 9 (nove) questões onde, algumas dessas, pedem mais do que um item para ser respondido, como no caso das questões de números 2, 3, 4, 5, 6 e 10. As questões 2, 3 e 6 apresentam três itens cada uma; as questões 5 e 10 apresentam 2 itens cada uma; a questão 4 apresenta

quatro item e as questões 7, 8 e 9 apresentam um item cada um, perfazendo um total de 20 itens.

A seguir, apresentaremos os objetivos do teste e sua construção.

### **3.2.2.1. O TESTE: SEUS OBJETIVOS E CONSTRUÇÃO**

O objetivo de nossa pesquisa é investigar quais são os conhecimentos básicos dos alunos dos cursos de Licenciatura em Matemática, Pedagogia e Administração sobre a leitura e interpretação de gráficos e tabelas estudados na disciplina de Estatística, para que seja possível fazer um estudo comparativo entre tais conhecimentos.

Nesse sentido elaboramos um teste que priorizasse a investigação dos conhecimentos desses sujeitos, focando especialmente na:

- a) **Leitura e interpretação de gráficos:** de colunas, de barras ou setor
- b) **Construção de gráficos:** de colunas, de barras ou setor
- c) **Leitura e interpretação de tabelas:** simples e dupla entrada
- d) **Construção de tabelas:** simples e dupla entrada
- e) **Conversão entre os registros de representações:** registro de tabelas, registro gráfico, registro de sistemas de escrita, registro da língua natural e as combinações entre si.

E em caráter secundário:

- f) **Cálculo de medidas de tendência central:** moda, mediana e média

Encontraremos um desenho dos objetivos de cada questão na Figura 25.

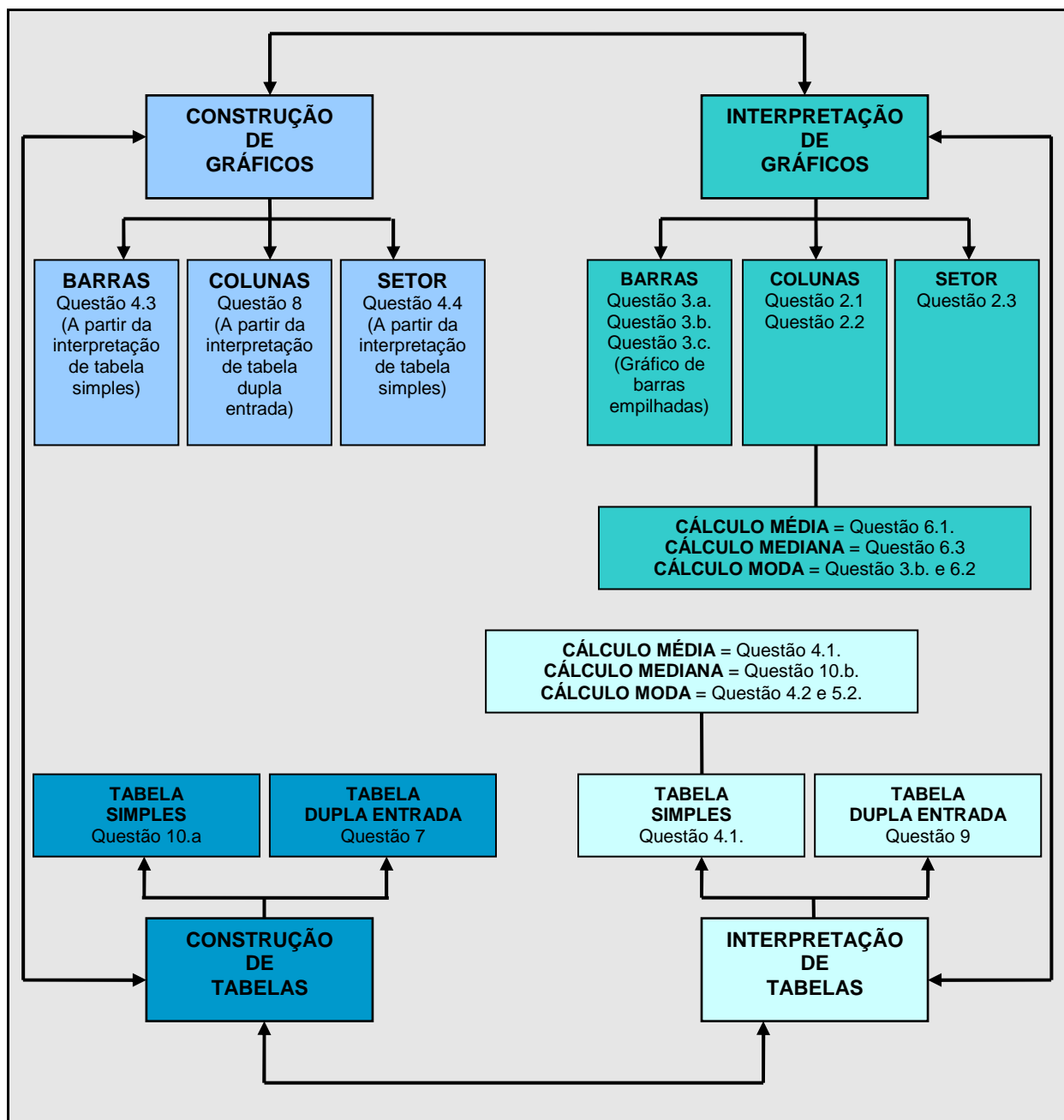


Figura 25: Desenho dos objetivos de cada questão.

Definido os objetivos a ser trabalhado no teste, partimos para a construção do questionário sobre o perfil dos sujeitos respondentes e o teste propriamente dito. Este foi submetido pelo menos três vezes ao crivo do grupo REPARE, onde seus membros opinaram e apontaram as possíveis correções. Essa opção de submeter o teste para a análise do grupo é uma forma de obter diferentes opiniões e vários olhares de especialistas na área de Educação Matemática, a fim de apurar nosso teste.

Logo após esse momento, o teste-piloto foi aplicado em 5 (cinco) professores voluntários, sendo 3 (três) professores Licenciados em Matemática e 2 (dois) Pedagogos que lecionam em uma escola pública do município de São Bernardo do Campo no Estado de São Paulo, escolhida por conveniência e de forma não aleatória. A finalidade dessa aplicação é a de obter um novo refinamento do teste e levantar possíveis falhas na interpretação das questões sobre a leitura e interpretação de gráficos e tabelas e o cálculo de medida de tendência central.

Nesse sentido e de posse das informações apuradas no teste-piloto, partimos para a aplicação do teste principal nas instituições escolhidas.

A seguir, encontraremos uma discussão sobre o tipo de análise que pretendemos realizar, os objetivos de cada questão e as respostas esperadas.

### **3.2.2.2. UMA ANÁLISE A PRIORI DO TESTE**

Com os dados colhidos por meio do teste pretendemos traçar duas análises, uma com abordagem quantitativa no que tange ao percentual de acertos e erros das questões e outra com abordagem qualitativa no que diz respeito às estratégias de resolução escolhidas pelos sujeitos a fim de obter duas visões complementares e traçar um comparativo entre os grupos sobre o mesmo ponto de estudo: o conhecimento básico dos alunos dos cursos de Licenciatura em Matemática (GFLM), Pedagogia (GFP) e Administração (GFADM) sobre a leitura e interpretação de gráficos e tabelas estudados na disciplina de Estatística.

Acreditamos que essa análise quantitativa e comparativa seja suficiente para refutar ou confirmar nossa hipótese de que os alunos do **GFADM** mobilizarão mais conhecimentos frente às situações que envolvem os conhecimentos básicos estatísticos como leitura e interpretação de gráficos e tabelas, e, conseqüentemente apresentarão melhor desempenho nos testes aplicados do que os alunos dos grupos **GFLM** e **GFP**. Acreditamos que isso possa ocorrer por que os gráficos, tabelas e as medidas de tendência central, possivelmente, façam mais sentido ao **GFADM** – uma

vez que as utilizam em outras disciplinas que não seja a de Estatística - do que para o grupo **GFLM** e do **GFP**.

Quanto à abordagem qualitativa de cada questão, pretendemos analisar as estratégias empregadas pelos sujeitos sob quatro perspectivas, a saber:

- a) Perspectiva dos níveis de compreensão gráfica de Curcio (1989): na questão de leitura dos dados, leitura entre os dados ou leitura além dos dados.
- b) Perspectiva das regras de construção de tabelas de Wainer (1992):
- c) Perspectiva dos conceitos elementares de Estatística com relação às regras de construção de gráficos e tabelas.
- d) Perspectiva da Teoria de Registros de Representação de Duval (2005): na questão das conversões entre os de registros de representação.

Sob o crivo dessas quatro perspectivas elaboramos nossas **unidades de análise** e suas respectivas **categorias**:

- a) **Unidade de Análise Construção de Gráficos** - onde iremos analisar se ao construir o gráfico o aluno:
  - Anotou as unidades utilizadas nas escalas;
  - Gráfico apresenta título ou legenda;
  - Quando gráfico de setor, este apresenta área proporcional às frequências de cada categoria
  - Em branco (não registrou nenhum tipo de gráfico).
- b) **Unidade de Análise Construção de Tabelas** – onde iremos analisar se a tabela construída pelo aluno:
  - Apresenta título;
  - O título responde as perguntas o quê, onde e quando;
  - Apresenta coluna indicadora ou cabeçalho;
  - A indicação da coluna indicadora ou cabeçalho é coeso com as informações da tabela;
  - O número de linhas ou colunas está correto;

- Apresenta fonte;
  - A estrutura está em ordem decrescente;
  - Apresenta valor total;
  - Em branco (não registrou nenhum tipo de tabela).
- c) **Unidade de Análise Leitura e Interpretação de Gráficos** – onde iremos analisar se o aluno se enquadra no:
- Nível “Leitura dos Dados”;
  - Nível “Leitura entre os Dados”;
  - Nível “Leitura além dos Dados”;
  - Em branco (não registrou nenhum tipo de estratégia de leitura ou interpretação de gráficos).
- d) **Unidade de Análise Leitura e Interpretação de Tabelas** - onde iremos analisar se o aluno se enquadra no:
- Nível básico;
  - Nível Intermediário;
  - Nível Avançado;
  - Em branco (não registrou nenhum tipo de estratégia de leitura ou interpretação de tabelas).
- e) **Unidade de Análise Representação Semiótica:** onde iremos analisar se o aluno
- Realiza a conversão de Registro Gráfico para Registro Tabular;
  - Realiza a conversão de Registro Tabular para Registro Gráfico;
  - Realiza a conversão de Registro Tabular para Registro Numérico (medidas de tendência central – moda, mediana e média);
  - Realiza a conversão de Registro Gráfico para Registro Numérico (medidas de tendência central – moda, mediana e média);
  - Realiza o tratamento dentro do Registro Gráfico.

E em caráter secundário:

- f) **Unidade de Análise Calculo de Moda Média e Mediana** – analisaremos o percentual de acerto e erro das questões que envolvem o cálculo da Moda, Média e Mediana.

Para tanto iremos traçar uma análise a priori das questões de número 2 até o número 10 do teste aplicado nos grupos GFLM, GFP e GFADM, excluindo a questão de número 1 que trata do perfil dos respondentes.

Questão 2.

2. Observe os gráficos abaixo de uma pesquisa sobre o turismo doméstico e responda:



a) Embratur.

b) Embratur.

Fonte: Projeto Araribá: Matemática/obra coletiva. São Paulo: Moderna, 2006.

2.1. Para saber a quantidade de turistas que se hospedou em hotel no ano de 2001, precisamos ler as informações de qual desses gráficos?

- Somente o gráfico (a)  
 Somente o gráfico (b)  
 Os dois, (a) e (b)  
 Nenhum dos dois

2.2. Qual a informação que o gráfico (a) traz?

- A quantidade de turistas que realizaram viagens no Brasil no ano de 1998  
 A quantidade de turistas que realizaram viagens no Brasil no ano de 2001  
 A quantidade de turistas que realizaram viagens no Brasil no ano de 1998 e 2001  
 A quantidade de turistas que realizaram viagens no Brasil no ano de 1998 a 2001

2.3. O gráfico (b) traz que tipo de informação?

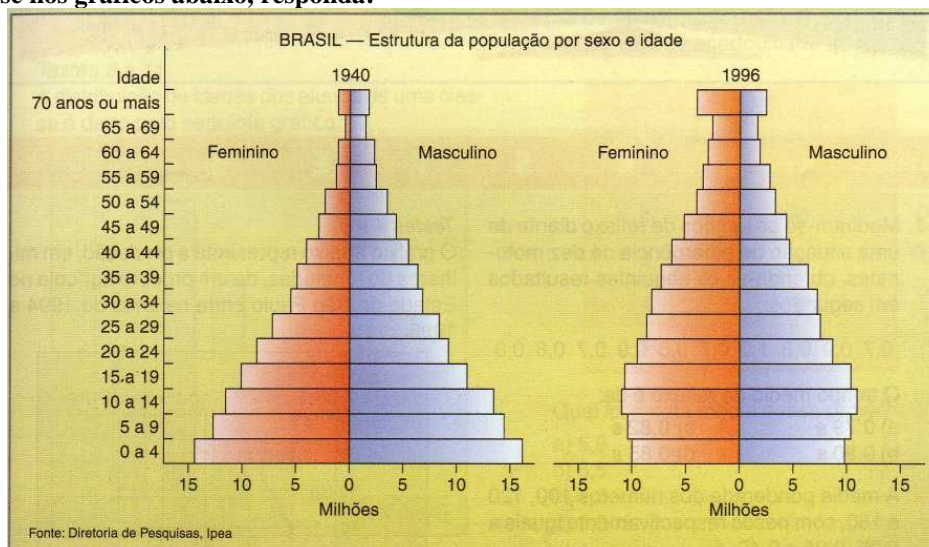
- O percentual de turistas de acordo com os tipos de transporte utilizado.  
 O número de turistas de acordo com o tipo de hospedagem utilizada.  
 O percentual de turistas de acordo com o tipo de hospedagem utilizada nos anos de 1998 e 2001.  
 O percentual de turistas de acordo com o tipo de hospedagem utilizada no ano de 1998.

A questão 2, de uma forma geral, aborda a interpretação de gráficos de coluna e de setor. Todos os itens propõem a localização de informações que estão explícitas nos dois gráficos e exige do leitor apenas “nível leitura dos dados”.

Nossa expectativa é a de que teremos altos índices de acertos nesta questão, uma vez que o grau de compreensão cognitivo exigido é muito baixo.

### Questão 3.

#### 3. Com base nos gráficos abaixo, responda:



Fonte: Iezzi, G. Dolce, O. Machado, A. - Matemática e Realidade: 7ª série - 4 ed. Reform. - São Paulo: Atual, 2000.

- Em 1940, na faixa etária de 0 a 4 anos, havia mais homens ou mais mulheres no Brasil? \_\_\_\_\_ E em 1996? \_\_\_\_\_
- Em que faixa estava a idade modal masculina em 1996? \_\_\_\_\_
- Qual a faixa etária que tem menos população em 1940? \_\_\_\_\_

A questão 3, também trabalha com a interpretação de informações explícitas nos gráficos do tipo “barras empilhadas” e traz a comparação entre os gêneros e também o cálculo da moda.

No item (a) solicitamos que os alunos comparem os gêneros e aponte qual deles estão em número maior, porém o aluno deve levar em consideração as cores dos gráficos (azul para masculino e vermelho para feminino), os anos que cada gráfico representa e observar bem a escala utilizada para a construção do eixo das frequências. Nossa expectativa com relação ao desempenho dos alunos é a de que teremos altos índices de acertos, por se tratar do nível “leitura dos dados” e sem muito esforço cognitivo.

No item (b), que trabalha com a medida de tendência central “moda”, encontramos uma questão no nível “leitura além dos dados”, uma vez que é solicitado que o sujeito lance mão de um conceito que não está explícito no gráfico, porém, não apresenta um nível elevado de dificuldade. A expectativa do número de acertos não é muito alta, por acreditar que os sujeitos terão dificuldades em localizar a idade modal no gráfico de “barras empilhadas”.

O item (c) trabalha com uma questão no nível “leitura dos dados” e que exige pouco esforço cognitivo, uma vez que o sujeito deve procurar uma informação que está explícita no gráfico. A expectativa do número de acertos desta questão é alta.

Questão 4.

Na tabela abaixo estão computadas as opiniões fictícias de 60 pessoas sobre um filme que acabava de estrear na cidade de Argolina.

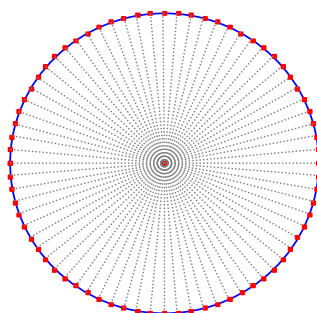
<b>OPINIÃO</b>	<b>Nº DE PESSOAS</b>
Excelente	9
Ótimo	15
Bom	18
Regular	12
Ruim	3
Péssimo	3
<b>TOTAL</b>	<b>60</b>

Fonte: Iezzi, G. Dolce, O. Machado, A. - Matemática e Realidade: 6ª série - São Paulo:Atual, 2000.

<p>4.1. É possível calcular a média dos dados acima?  <input type="checkbox"/> Sim   <input type="checkbox"/> Não                  Por quê?</p>	<p>4.2. Indique a opinião mais frequente (moda).</p>
---	--

4.3 Represente os dados da tabela em um gráfico

4.4. Calcule as porcentagens relativas às opiniões e represente-as num gráfico de setores. Para esta construção utilize a circunferência abaixo que já está dividida de 5° em 5°.



Os itens de número 4.2, 4.3 e 4.4 foram elaborados pensando nas possíveis mudanças de registros do tipo “conversão”, enquanto que o item 4.1 exige a leitura tabular no “nível intermediário”.

No item 4.1 é solicitado que os sujeitos analisem a possibilidade de efetuar ou não o cálculo da medida de tendência central “média” de variáveis qualitativas e que disserte sobre sua escolha. Esta é uma questão que não admite a mudança do

registro de tabela para o registro numérico (média) justamente por se tratar de uma variável qualitativa nominal, fato este que torna esta questão da unidade “leitura e interpretação de tabelas”. A expectativa do número de acertos desta questão é baixa. Acreditamos que os sujeitos não levarão em conta o fato de estarem trabalhando com uma variável qualitativa nominal (que não é passível de outro cálculo de medida de tendência central que não a Moda) e irão calcular erroneamente a média entre o número de frequências que cada uma das variáveis apresenta.

O item 4.2 aborda uma questão que exige um esforço cognitivo médio, uma vez que precisamos localizar uma informação que está explícita na tabela, porém o sujeito deve fazer a opção pela “opinião” em vez do “nº de pessoas”. Esperamos um bom índice de acertos nesta questão.

Já na questão 4.3, temos um caso de mudança de registro de representação do tipo conversão. Os dados estão registrados em uma tabela e os sujeitos devem convertê-la para a representação gráfica. Trata-se de uma tabela simples com variáveis qualitativas nominais e esperamos que os sujeitos construam um gráfico de barras ou colunas justamente por que esse tipo de gráfico pode acomodar um número grande de categorias em qualquer nível de mensuração. Esperamos um alto índice de acertos quanto à mudança de registro, porém quando analisaremos qualitativamente a construção, acreditamos que os sujeitos não terão os cuidados necessários quanto aos elementos de construção de um gráfico (escala, identificação das variáveis, construção das colunas, etc), logo baixos índices de acerto.

A questão 4.4 aborda um tratamento de dados registrados em gráfico de barras para a representação gráfica de setor, além da conversão do registro numérico para os registros de porcentagem e de graus. De acordo com o gráfico de variáveis nominais já construídos por eles, esperamos que os estes calculem a porcentagem do número de pessoas em cada uma das variáveis e converta-as para graus, com isso construa o gráfico de setor. Estamos falando de uma conversão tripla (registro gráfico para registro numérico de porcentagem depois para registro numérico em graus) e um tratamento dentro do registro gráfico. Esperamos um baixo índice de acerto nesta questão, por serem necessário muitos “passos” de conversão para chegar à resolução.

### Questão 5.

**5. Uma prova contendo cinco questões, cada uma valendo 2,0 pontos, foi aplicada numa classe de 40 alunos. Na tabela está o número de alunos que acertou cada questão:**

<i>Questão</i>	<i>Número de alunos que acertaram</i>
1	12
2	40
3	24
4	32
5	36

Fonte: Iezzi, G. Dolce, O. Machado, A. - Matemática e Realidade. São Paulo:Atual, 2000.

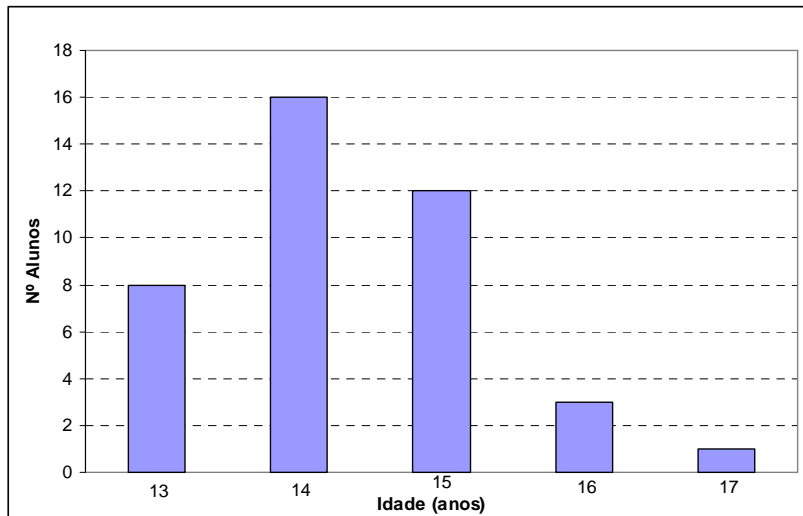
5.1. Onde houve maior variação do número de acertos?	5.2. Qual é a moda da questão mais respondida?
--	--

A questão de número 5 aborda uma tabela de distribuição de frequências simples com variáveis qualitativas nominais. No item 5.1, foi solicitado que os sujeitos encontrem onde ocorreu maior variação do número de acertos. Este é um item que foi elaborado pensando no “nível intermediário” uma vez que o sujeito irá efetuar uma comparação entre as variáveis registradas em uma tabela. A expectativa é alta em relação ao número de acertos, porém sempre observando se o sujeito indicará que esta variação ocorreu da “questão 1” para a “questão 2” e não somente indicar “questão 2”.

Para o item 5.2 elaboramos uma questão que não exige muito esforço cognitivo por parte do sujeito. Trata-se de uma questão de mudança de registro do tipo conversão do registro de tabela para o registro numérico “moda”. Além disso, enquadramos esta questão no “nível básico” da unidade Leitura e Interpretação de Tabelas. A expectativa em relação a esta questão também é alta.

## Questão 6

6. A distribuição de idades dos alunos de uma classe é dada pelo seguinte gráfico:



Fonte: Iezzi, G. Dolce, O. Machado, A. - Matemática e Realidade. São Paulo:Atual, 2000.

6.1. Qual é a idade média dos alunos?

6.2. A idade modal é:

6.3. Qual é a idade que representa a mediana?

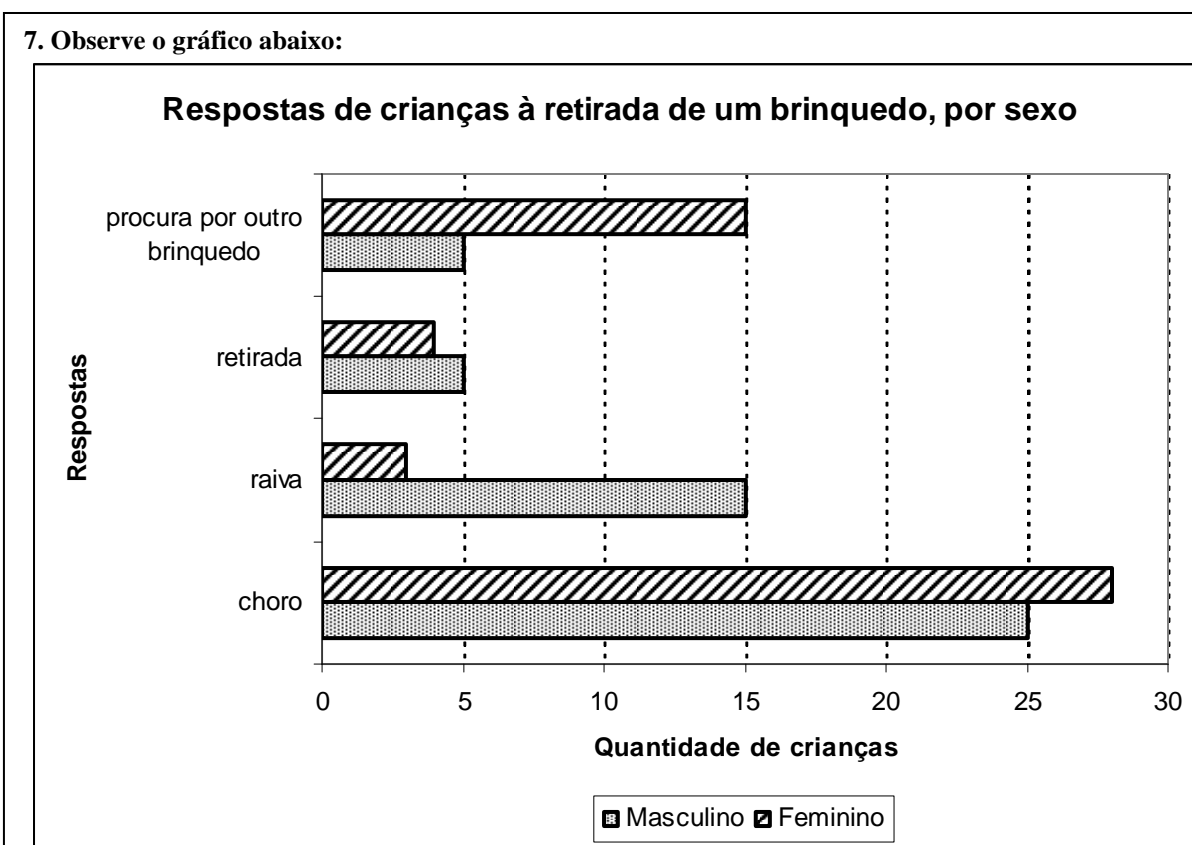
Com relação à questão 6, todos os itens foram elaborados com vistas na Leitura e Interpretação de Gráficos e, a questão 6.3 também na mudança de registro do tipo conversão, do registro gráfico para o registro numérico medida de tendência central.

Na questão 6.1 (leitura além dos dados) espera-se que o aluno calcule a idade média de dados representados em um gráfico de barras e apesar de não significar um cálculo muito elaborado e que por vezes é até intuitivo, esperamos um baixo índice de acertos pelo motivo de que os sujeitos tendem a não observar o número de vezes que as idades (anos) ocorreram.

A questão 6.2 também não exige um esforço cognitivo muito alto, os sujeitos devem localizar no gráfico a idade modal. Este item trabalha com uma questão no nível “leitura dos dados”, uma vez que a informação está explícita no gráfico e a expectativa do número de acertos é alta.

Já na questão 6.3, iremos analisar as respostas sob duas perspectivas: a conversão do Registro gráfico para o registro numérico (medidas de tendência central) e a leitura e interpretação de gráficos “leitura entre os dados”. Esperamos que os sujeitos calculem a mediana a partir de informações retiradas de um gráfico de barras. O índice de acerto esperado para essa questão não é muito positivo, uma vez que, acreditamos que os sujeitos não levarão em conta que o cálculo é uma média aritmética de dados agrupados, e, por esse motivo, irão somente localizar a idade central.

### Questão 7



Fonte: Adaptado de Levin, Jack. Fox, J.A; Estatística para ciência Humanas. São Paulo: Pearson Prentice Hall, 2004.

Agora com base na leitura dos dados contidos no gráfico de barras acima, construa uma tabela. Para facilitar a construção tem-se abaixo uma grade. Utilize o quanto for necessário e risque o restante.


Já a questão de número 7 apresenta uma atividade que solicita a conversão do registro gráfico para o registro de tabela e de forma qualitativa iremos analisar a construção da tabela.

Nessa questão temos um gráfico de barras horizontais que mostra o efeito de uma variável sobre a outra e esperamos que os sujeitos construam uma tabela de dupla entrada, uma vez que, com base nos dados do gráfico, trata-se de uma distribuição de idades por sexo. A expectativa para o número de acertos desta questão, na unidade conversão, é alta, uma vez que acreditamos que os sujeitos respondentes não sentirão dificuldades em expressar uma tabela de dupla entrada.

Já quando analisaremos qualitativamente a construção, acreditamos que os sujeitos não terão os cuidados necessários quanto aos elementos de construção de um gráfico (escala, identificação das variáveis, construção das colunas, etc), logo baixos índices de acerto.

#### Questão 8.

##### 8. Analise as informações da tabela:

**TABELA 2.15** Tabulação cruzada do uso do cinto de segurança por sexo

Uso do cinto de segurança	Sexo do entrevistado		Total
	Masculino	Feminino	
Sempre	144	355	499
Na maioria das vezes	66	110	176
Algumas vezes	58	66	124
Raramente	39	44	83
Nunca	60	55	115
Total	367	630	997

Fonte: Adaptado de Levin, Jack. Fox, J.A; Estatística para ciência Humanas. São Paulo: Pearson Prentice Hall, 2004.

Agora com base na leitura da tabela acima construa um gráfico, por sexo, do entrevistado:

Na questão 8 apresentamos a conversão de dados registrados em tabela de dupla entrada para o registro gráfico. De acordo com a tabela de variáveis nominais, esperamos que os sujeitos construam o gráfico de barras ou colunas que mostram o efeito de uma variável sobre a outra, em outros termos, que eles construam o gráfico

separado de duas em duas colunas (uma para masculino e outra para feminino) de cada variável. A expectativa, com relação ao número de acertos é alta, uma vez que não se trata de uma questão de alto nível cognitivo. No caso desta questão também analisaremos qualitativamente a construção do gráfico e todos seus elementos.

### Questão 9

9. (ENEM/2001) A tabela apresenta a taxa de desemprego dos jovens entre 15 e 24 anos estratificada com base em diferentes categorias.

Região	Homens	Mulheres
Norte	15,3	23,8
Nordeste	10,7	18,8
Centro-oeste	13,3	20,6
Sul	11,6	19,4
Sudeste	16,9	25,7
<b>Grau de Instrução</b>		
Menos de 1 ano	7,4	16,1
De 1 a 3 anos	8,9	16,4
De 4 a 7 anos	15,1	22,8
De 8 a 10 anos	17,8	27,8
De 11 a 14 anos	12,6	19,6
Mais de 15 anos	11,0	7,3

Fonte: PNAD/IBGE, 1998.






Considerando **apenas** os dados acima e analisando as características de candidatos a emprego, é possível concluir que teriam **menor** chance de consegui-lo,

- (A) Mulheres, concluintes do ensino médio, moradoras da cidade de São Paulo.
- (B) Mulheres, concluintes de curso superior, moradoras da cidade do Rio de Janeiro.
- (C) Homens, com curso de pós-graduação, moradores de Manaus.
- (D) Homens, com dois anos do ensino fundamental, moradores de Recife.
- (E) Mulheres, com ensino médio incompleto, moradoras de Belo Horizonte.

Na questão de número 9, trabalhamos com a leitura e interpretação de tabelas no “nível avançado”, uma vez que é necessário o sujeito ter noção de região geográfica para reconhecer a localização de algumas capitais brasileiras em relação à região. Além disso, os sujeitos também devem contabilizar o número de anos de estudo necessários para localizar o grau de instrução. Esperamos um baixo índice de acerto, justamente por ser uma questão que mobiliza conhecimentos além da Estatística e da Matemática.

### Questão 10

10. Na sala de aula de 4ª série J, a professora de Matemática solicitou que cinco alunos se colocassem em pé na frente da sala, medindo com uma fita métrica a estatura de cada um e anotando na lousa. Veja a estatura dos cinco alunos.

	Luiz 152 cm		Ana 148 cm		João 155 cm		
		Bia 145 cm				Caio 150 cm	

Adaptado de Carzola e Santana, Tratamento da Informação para o Ensino Fundamental e Médio. Bahia, Via Litterarum 2006.

a) Com base na leitura dos dados acima, construa uma tabela. Para facilitar a construção tem-se abaixo uma grade. Utilize o quanto for necessário e risque o restante.


b) Agora de acordo com a tabela que você construiu, calcule a altura mediana dos meninos da 4ª série J.

A questão 10 se divide em dois itens. No item (a) trabalhamos com a mudança de registro do tipo conversão de registro figural e numérico para registro de tabela. Também não é uma questão de alto nível cognitivo, por este motivo, esperamos um número elevado de acertos, porém com dificuldades em organizar a construção da tabela (análise qualitativa).

No item (b) é solicitado que o sujeito, baseado no registro figural, calcule a mediana dos dados. Trata-se de uma questão de mudança de registro do tipo conversão – do registro figural para o registro numérico (mediana). A expectativa com relação ao índice de acertos é alta, uma vez que é uma questão de nível cognitivo baixo. Também é importante ressaltar que, consideraremos a questão como certa, tanto para os sujeitos que calcularem a altura mediana só dos meninos

(levando em consideração o gênero) com também para aqueles que calcularem a altura mediana de todos os meninos (conotação de crianças). Essa medida foi tomada pelo fato de termos solicitado que os sujeitos calculassem “a altura mediana dos **meninos** da 4ª série J”, e a palavra “meninos” dá interpretação ambígua.

### ANALISANDO OS RESULTADOS OBTIDOS

---

No capítulo anterior apresentamos do ponto de vista teórico, a metodologia escolhida, o delineamento do universo de estudo, os sujeitos, o material de estudo, o objetivo e a construção do teste, bem como uma análise a priori das questões arroladas na pesquisa.

Este capítulo é destinado à análise dos resultados obtidos de nosso instrumento diagnóstico, aplicado para 72 (setenta e dois) alunos de Licenciatura em Matemática, 48 (quarenta e oito) alunos de Pedagogia e 54 (cinquenta e quatro) alunos de bacharelado em Administração. Tal análise será dividida em duas etapas:

- \* Primeira etapa: focaremos na análise quantitativa do desempenho geral de cada grupo (GFLM, GFP, GFADM). Faremos isso por meio das taxas percentuais de acertos desses grupos. Nessa etapa ainda procederemos com a análise qualitativa, a partir de uma discussão sobre os tipos de erros cometidos pelos grupos.
- \* Segunda etapa: de caráter secundário, uma vez que não é foco de nossa pesquisa, apresentaremos uma análise quantitativa das questões em que era solicitado o cálculo de medidas de tendência central (moda, média e mediana).

#### **4.1. ANÁLISE QUANTITATIVA E QUALITATIVA: O DESEMPENHO TRADUZIDO EM NÚMEROS**

Essa análise pretende traçar um comparativo entre os desempenhos dos grupos no teste aplicado. Aqui o que importa é o acerto ou erro na questão. As respostas em branco foram computadas como erradas.

Para fazer a comparação entre os desempenhos dos grupos, realizamos um tratamento estatístico baseada nos testes: teste  $F$  (análise de variância), teste de Tukey (DHS) e qui-quadrado ( $\chi^2$ ).

Três focos serão abordados na análise: 1) a leitura e interpretação adequada de gráficos; 2) a leitura e interpretação adequada de tabelas; 3) a representação semiótica em Estatística, do tipo conversão e tratamento. Chamamos esses focos de *unidade de análise* (já discutida no capítulo III).

O Quadro 5 relaciona as três unidades de análise com cada uma das categorias e as suas respectivas questões.

<b>Unidades de Análise</b>	<b>Categorias</b>	<b>Itens</b>
Leitura e Interpretação de Gráficos	Nível "Leitura dos Dados"	2.1.a; 2.2.c; 2.3.c; 3.a.a; 3.a.b; 3.c e 6.2
	Nível "Leitura entre os Dados"	6.3 e 3.b
	Nível "Leitura além dos Dados"	6.1
Leitura e Interpretação de Tabelas	Nível Básico	5.2
	Nível Intermediário	4.1.a, 4.1.b e 4.2
	Nível Avançado	5.1 e 9.e
Representação Semiótica em Estatística	Conversão do Registro Gráfico para o Registro Tabular	7 e 10.0
	Conversão do Registro Tabular para Registro Gráfico	4.3 e 8
	Conversão do Registro Gráfico para o Registro Numérico (medidas de tendência central)	6.3
	Conversão do Registro Tabular para o Registro Numérico (medidas de tendência central)	5.2 e 10.b
	Tratamento dentro do Registro Gráfico	4.4

Quadro 5: Distribuição dos itens em relação às unidades de análise e suas categorias.

Cabe aqui lembrar que a unidade de análise "Leitura e interpretação de Gráficos" apresenta suas categorias de acordo com os níveis de compreensão gráfica de Curcio (1989); a unidade de análise "Leitura e interpretação de Tabelas"

com os níveis de representação tabular de Wainer (1992) e “Representação Semiótica em Estatística” com a teoria de Duval (2005).

Então, para a análise quantitativa, consideraremos 24 (vinte e quatro) itens repondidos por 174 (cento e setenta e quatro) sujeitos, perfazendo um total de no máximo 4176 possíveis respostas corretas, divididas em:

- GFLM com 72 sujeitos respondendo 24 itens, o que significa a possibilidade de obtermos 1728 respostas corretas;
- GFP com 48 sujeitos respondendo 24 itens, implicando em 1152 possíveis respostas corretas;
- GFADM com 54 sujeitos respondendo 24 itens, o que gera a possibilidade de 1296 respostas corretas.

#### **4.1.1. A ANÁLISE GERAL DOS GRUPOS**

A análise do desempenho geral dos grupos (quantitativa) será seguida pela análise dos tipos de erro cometidos por esses grupos (qualitativa). Acreditamos que ao proceder com uma análise quantitativa e qualitativa, teremos uma visão mais completa dos grupos, o que, por sua vez, permitirá traçar um panorama mais acurado da situação destes três grupos de alunos em relação aos conhecimentos sobre leitura e interpretação de gráficos e tabelas.

O gráfico da Figura 26 apresenta as médias gerais obtidas pelos grupos.

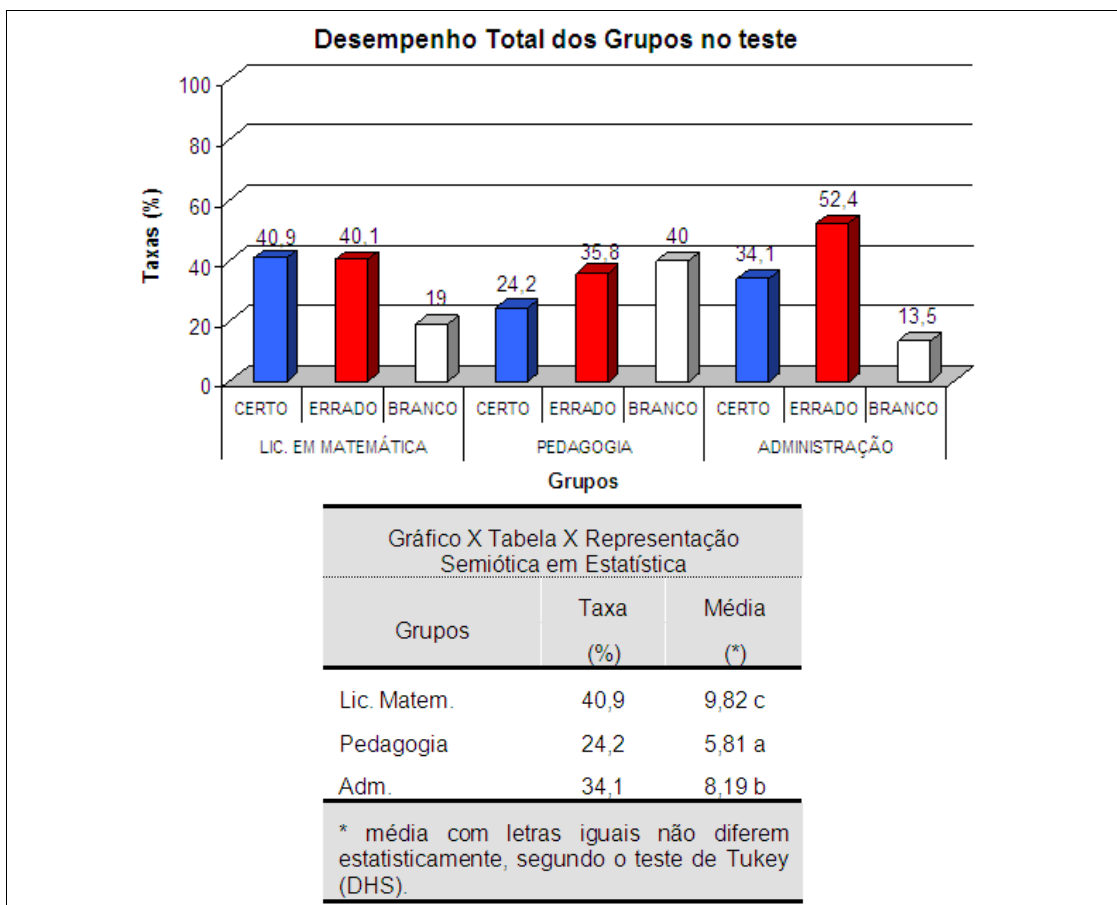


Figura 26: Desempenho total dos grupos no teste e tabela com respectivo teste estatístico de Tukey.

O resultado do teste de Tukey, comprova que os desempenhos dos grupos diferem significativamente [ $F(2,173) = 32,038$ ;  $p = 0,000$ ] uns dos outros, com o grupo dos estudantes de Pedagogia apresentando o menor sucesso e os de Licenciatura em Matemática o maior.

Queremos chamar a atenção para o alto índice de questões deixadas em branco pelo GFP. Observamos que esta taxa (40%) é praticamente equivalente à taxa de acerto do GFLM e, ainda, superior à taxa de acerto e branco do GFADM.

Esse dado nos leva a conjecturar que esses futuros Pedagogos analisados, possivelmente não apresentam uma atitude positiva em relação à Estatística, uma vez que não procuraram resolver as questões que necessitavam de cálculos. Apenas resolviam as de múltipla escolha que abordavam leitura e interpretação de gráficos no nível “leitura dos dados” e que não exigiam grande esforço cognitivo.

Esse alto percentual de questões deixadas em branco pelo GFP corrobora com as ideias de Grácio, Oliveira e Oliveira (s.d.) quando apontam para a falta de

motivação de alguns alunos de cursos superiores, entre eles, os de Pedagogia. O exemplo de resposta de um sujeito de Pedagogia da Figura 27, apoia nossa conjectura.

1.15. Você considera o Ensino de Estatística:

- ( ) muito importante
- ( ) importante
- ( ) pouco importante
- ( ) não é importante
- (✓) não sei avaliar

1.16. Dê três razões de sua escolha na questão 1.15

Pode ser que sim e pode ser que não.

Eu não pretendo usá-lo

Figura 27: Exemplo de resposta de sujeitos de Pedagogia para os itens de perfil 1.15 e 1.16.

Quanto aos resultados dos itens errados, observamos que o GFADM detém o maior índice (52,4%) em relação aos outros dois grupos, no entanto se considerarmos que esse foi o grupo que apresentou o menor percentual de resposta em branco, podemos supor que os estudantes deste grupo erravam na tentativa de acertar a questão.

Na Figura 28 apresentamos um gráfico – o boxplot – com os resultados positivos obtidos por cada um dos grupos. Esses resultados estão apresentados em função do número de acertos dos grupos (e não do percentual) no instrumento diagnóstico.

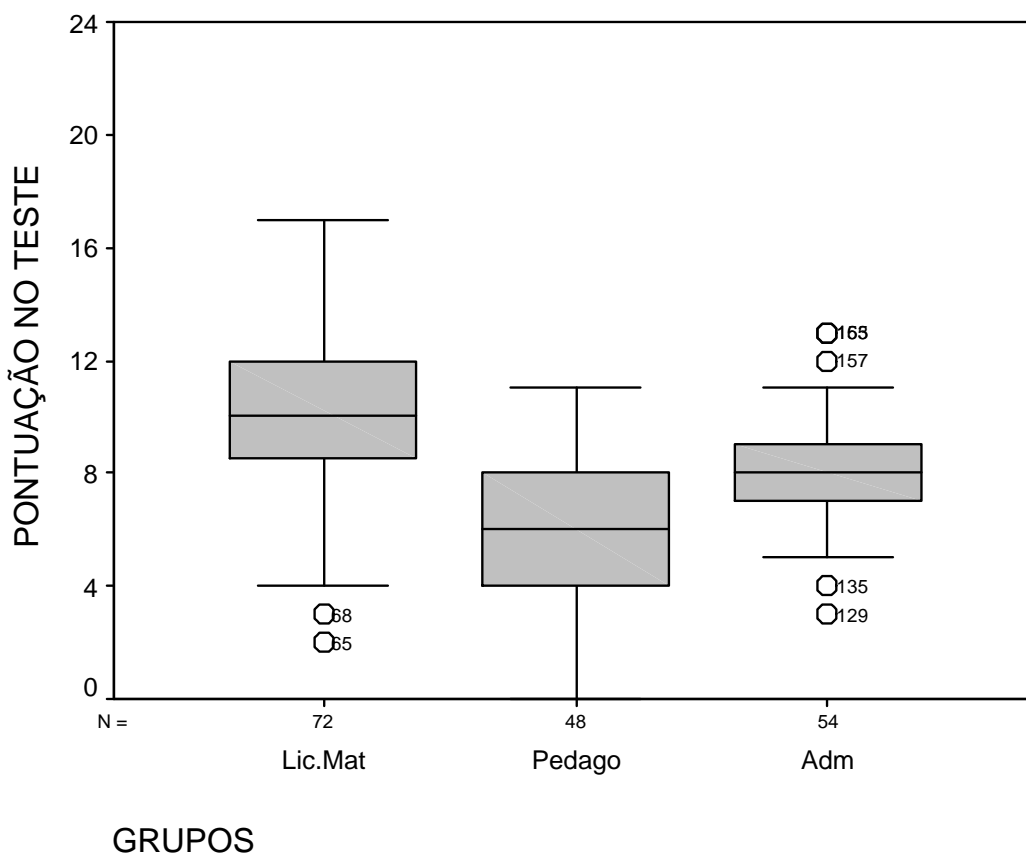


Figura 28: Boxplot com o desempenho dos grupos no teste.

Os resultados apurados dos itens corretos deixam claro que o desempenho do GFLM (com uma média de 9,82) foi mais positivo do que o desempenho do GFADM (com média de 8,19) e, quando comparado ao GFP (com média de 5,81) percebemos uma diferença consistente, que se justifica estatisticamente.

Apesar de nenhum dos grupos alcançarem a taxa de pelo menos 50% de acerto no teste, o fato do desempenho dos GFLM e GFADM ser mais positivo do que o GFP, supomos que pode ser atribuído ao maior tempo de estudo dedicado à disciplina de Estatística em seus cursos do que o de Pedagogia. Observe a Figura 29, que traz a resposta de um sujeito de Pedagogia que ampara nossa suposição.

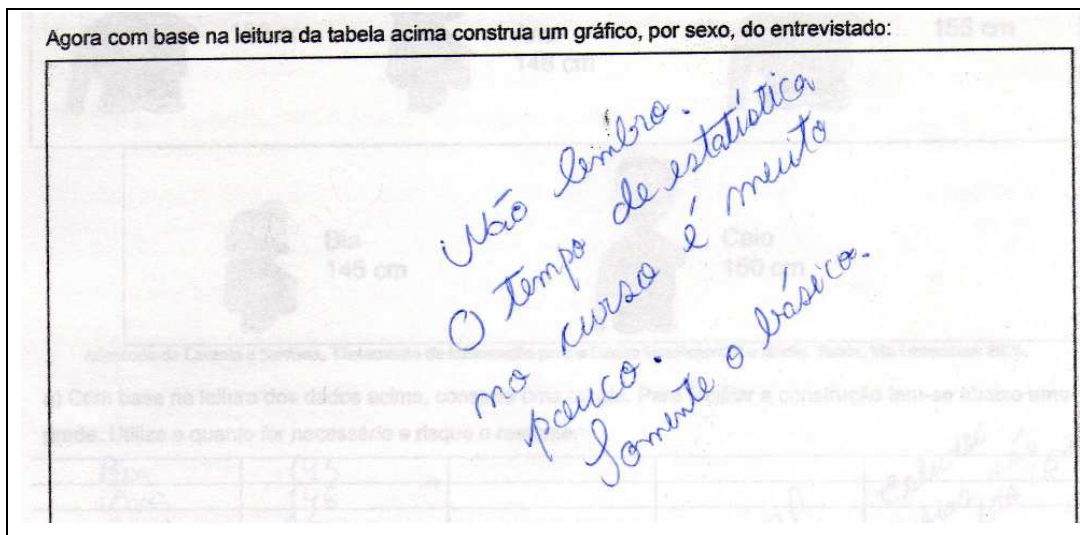


Figura 29: : Exemplo de respostas dos sujeitos de Pedagogia quanto ao tempo de estudo na disciplina Estatística.

#### 4.1.2. A ANÁLISE QUANTITATIVA DO DESEMPENHO POR UNIDADE DE ANÁLISE

Como nosso teste não apresenta número igual de questões em cada uma das categorias, atribuímos uma pontuação no teste que podia variar de 0 a 24 pontos. Em seguida, padronizamos essa pontuação em uma escala de zero a 10, multiplicando a pontuação de cada categoria por dez e dividindo-a pelo número de variáveis daquela categoria em questão, denominado por *nota* esse resultado.

A análise quantitativa do desempenho por unidade de análise permitirá responder algumas das questões de caráter mais específicos levantadas no Capítulo I como: “Qual é o nível de compreensão gráfica que o grupo dos futuros licenciados em Matemática (GFLM), o grupo dos futuros Pedagogos (GFP) e o grupo dos futuros Administradores (GFADM) apresentam?”; “Qual é o nível de representação tabular, que o GFLM, o GFP e o GFADM apresentam?”; “Qual grupo tem melhor desempenho na conversão de dados estatísticos apresentados na forma de tabelas para gráficos e vice-versa?” e, de forma secundária, a questão “Quais dos grupos se saem melhor em questões que envolvem cálculos de medidas de tendência central: o GFLM, o GFP ou o GFADM?”

Para tanto, iremos primeiro analisar o desempenho dos grupos de uma forma geral, dentro de cada uma das unidades de análise.

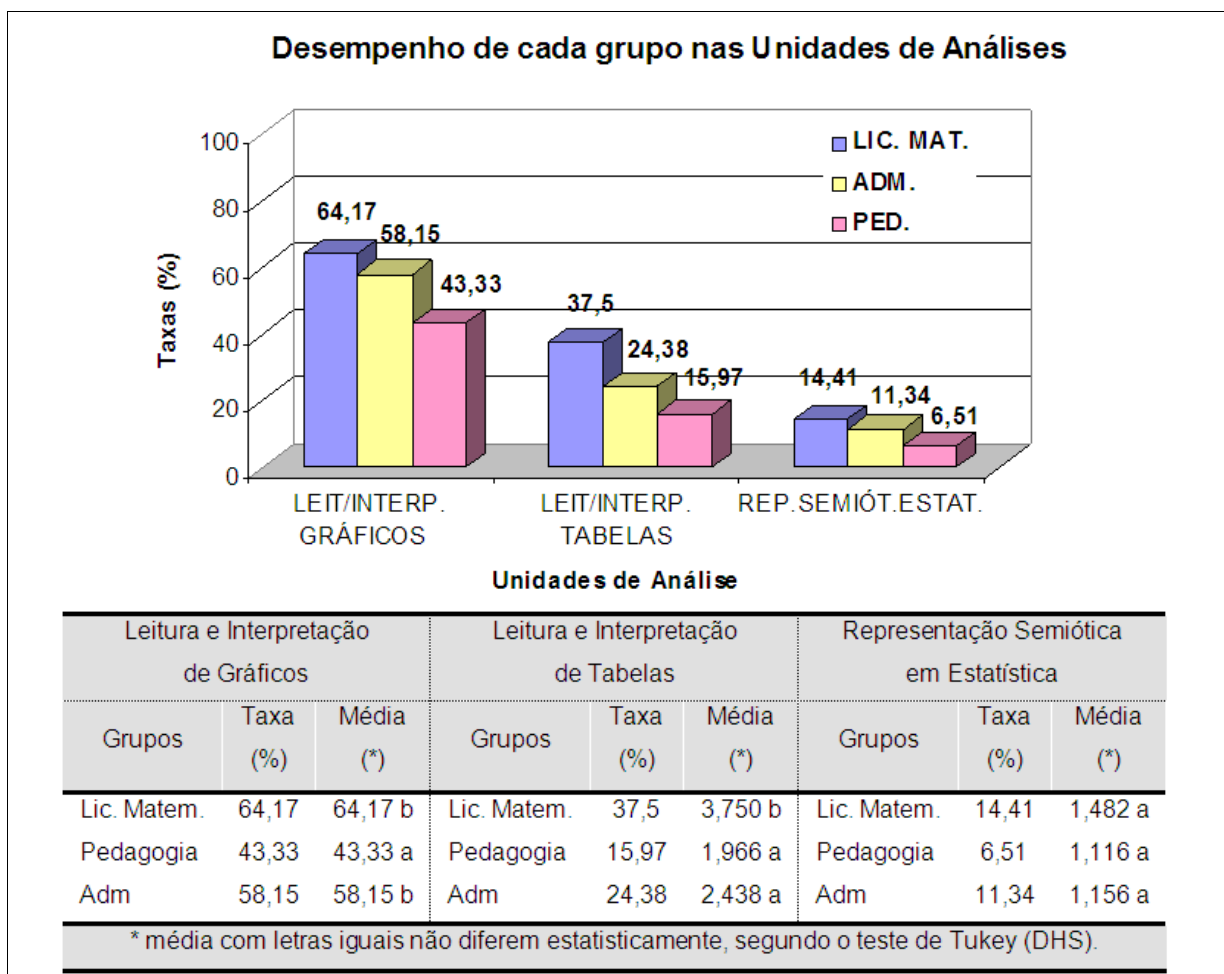


Figura 30: Desempenho total dos grupos em cada Unidade de Análise e tabela com os respectivos testes de Tukey.

A primeira observação que o gráfico da Figura 30 nos permite fazer é que o comportamento dos três grupos segue a mesma tendência, com as questões referentes à leitura e interpretação de gráficos apresentando o maior percentual de acerto e as questões classificadas como “representação semiótica” aquelas em que os sujeitos dos grupos se saíram pior. No entanto, embora seja clara a tendência similar nos três grupos, seus desempenhos diferem significativamente tanto nas questões de Leitura e Interpretação de Gráficos [ $F(2,171) = 24,220$ ;  $p = 0,000$ ] quanto na Leitura e Interpretação de Tabelas [ $F(2,160) = 7,493$ ;  $p = 0,001$ ]. Já no que tange às questões de Representação Semiótica em Estatística, não houve diferença estatisticamente significativa [ $F(2,148) = 1,658$ ;  $p = 0,194$ ].

A segunda observação é que o gráfico também mostra que os grupos foram bem mais sucedidos nas questões que abordavam a Leitura e Interpretação de Gráficos, seguidos da Leitura e Interpretação de Tabelas e, logo após, das questões que abordavam a Representação Semiótica em Estatística.

Quando analisamos a unidade *Leitura e Interpretação de Gráficos*, podemos observar que, apesar de serem os melhores índices obtidos pelos três grupos, eles não passam de cerca de 70%. Segundo Curcio (1989), os gráficos representam um suporte de comunicação de dados que já foram coletados, organizados e analisados e necessitam de uma leitura capaz de interpretar tais informações. Além disso, para Duval (1999, apud FLORES e MORETTI, 2005) as representações gráficas desencadeiam quatro funções cognitivas do pensamento: a função de comunicação, de tratamento, de objetivação e de identificação. Nesse sentido, podemos apontar que os sujeitos pesquisados obtêm um desempenho positivo quando pensamos na função de comunicação e identificação dos gráficos, elementos mais trabalhados na vida acadêmica em detrimento do tratamento e objetivação.

Quando passamos a olhar para as funções de tratamento e objetivação, os desempenhos se tornam menos positivos. O fato dos índices de acerto na unidade *Leitura e Interpretação de Tabelas* serem menor do que o de Gráficos, só corrobora com as ideias de Duval (2002, apud ARAÚJO e FLORES, 2005) quando alerta que nem sempre ler e interpretar uma tabela são tarefas tão simples.

Os baixos índices de acerto na unidade *Representação Semiótica em Estatística*, levam-nos a crer que estes alunos ainda não desenvolveram a apreensão conceitual dos objetos matemáticos (*noésis*) e conseqüentemente não coordenam os vários tipos de representação (*semiósis*).

Partiremos assim para a análise do desempenho dos três grupos em cada uma das unidades de análise e suas categorias, a saber:

#### **4.1.2.1. LEITURA E INTERPRETAÇÃO DE GRÁFICOS E SUAS CATEGORIAS**

Essa unidade de análise conta com 10 itens divididos nas seguintes categorias: 7 abordam o nível de compreensão gráfica nível “leitura dos dados”, 2

itens abordam o nível “leitura entre os dados” e 1 item no nível “leitura além dos dados”.

Nesse momento, iremos apresentar uma análise do desempenho dos sujeitos dentro de cada unidade e logo após, uma análise por categorias (níveis de leitura).

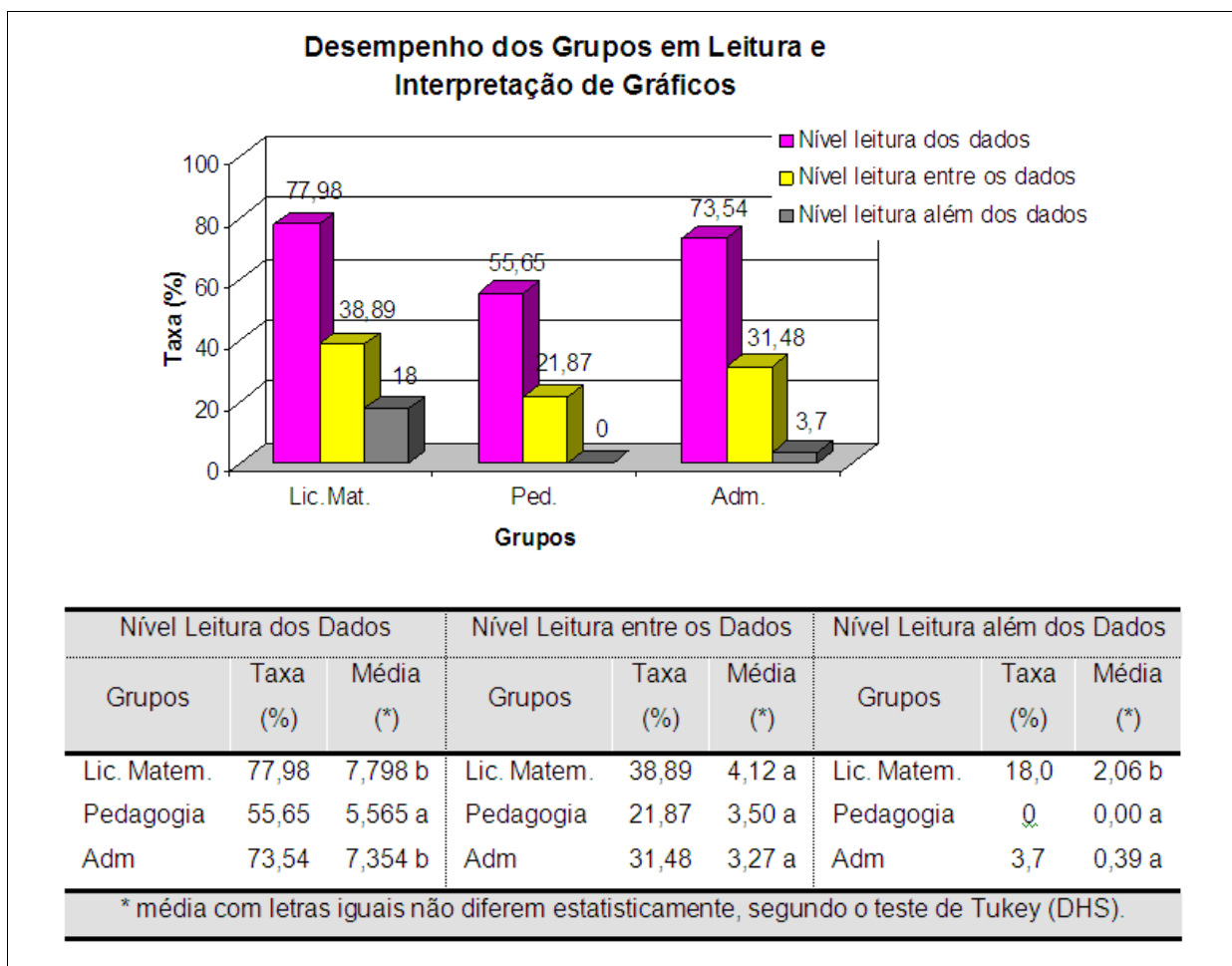


Figura 31: Desempenho dos sujeitos, segundo os grupos, nas categorias de Leitura e Interpretação de Gráficos e tabela com respectivos testes de Tukey.

Quando procedemos uma comparação entre os desempenhos dos grupos, vemos que seus comportamentos diferiram significativamente nos níveis “leitura dos dados” e “leitura além dos dados”, [ $F(2, 173) = 22,828$ ;  $p = 0,000$ ] e [ $F(2, 138) = 6,346$ ;  $p = 0,002$ ] respectivamente. Ao aplicarmos o teste de Tukey (DHS) constatamos que o desempenho do grupo dos Pedagogos nas questões “leitura dos dados” foi significativamente menos positivo que os apresentados pelos dos grupos de Administração e Licenciatura em Matemática. Já com relação ao nível “leitura além dos dados” não houve diferença significativa entre os desempenhos dos

grupos de Pedagogia e de Administração, mas esses foram significativamente menos positivos do que o desempenho do grupo de Licenciatura em Matemática. Por fim constatamos que no que tange ao nível “leitura entre os dados” não houve diferença significativa (segundo os testes *F* e Tukey) entre os desempenhos dos três grupos.

Observamos que o gráfico da Figura 31 está em concordância com a teoria de Curcio (1989). Conforme os níveis de leitura vão se elevando, as dificuldades se tornam maiores e os desempenhos, menores.

Vemos que os três grupos foram bem mais sucedidos nas questões que abordavam o nível “leitura dos dados”, apesar dos futuros Pedagogos acertarem pouco mais de 50% das questões. Quanto à análise a priori, já esperávamos que os sujeitos obtivessem bom desempenho nessa categoria. Tal fato corrobora com as ideias de Curcio (1989), o qual argumenta que esse é um nível onde o leitor apenas compreende os fatos explícitos, não existindo, assim, interpretação, o que demonstra ser de baixo custo cognitivo.

Em conformidade com as ideias de Curcio (1989), a medida que os níveis de leitura vão se elevando as dificuldades se tornam maiores. Nesse sentido, nossos resultados encontram-se em perfeita consonância com Curcio, já que os três grupos apresentam marcante decréscimo em seus percentuais de sucesso entre o nível mais simples – leitura dos dados – e o nível mais complexo – leitura além dos dados.

Para o nível “leitura entre os dados”, os desempenhos dos sujeitos se apresentam menos positivo do que o do nível anterior. Levantamos a hipótese que o baixo desempenho dos três grupos, nesse nível, aconteceu porque são itens que exigem cálculos integrados com base nas informações contidas no gráfico, o que exigiam uma leitura mais elaborada dos dados.

Já nos itens que abordavam a “leitura além dos dados”, os desempenhos são sofríveis, tanto para os futuros Licenciados em Matemática como para os Administradores, quanto aos futuros Pedagogos eles nem pontuaram o que conclui que esse nível é mais complexo que os dois anteriores, em que o leitor não encontra explícito no gráfico a resolução do problema, exigindo um alto custo cognitivo.

O resultado sobre as diferenças e similaridades nos comportamentos dos grupos no impelem a querer debruçar sobre as três categorias propostas por Curcio (“leitura dos dados”, “leitura entre os dados” e “além dos dados”) na busca de possíveis respostas para tais desempenhos.

### a) Categoria: Leitura dos dados

Essa categoria contava com 7 itens: Q.2.1, Q.2.2, Q.2.3, Q.3.a.a, Q.3.a.b, Q.3.c e Q.6.2. Desses, discutiremos os itens que apresentam uma diferença significativa, segundo o teste qui-quadrado.

Antes dessa discussão, promoveremos uma explanação à cerca dos resultados obtidos pelos sujeitos no item Q.2.1, por se tratar de um item de baixo custo cognitivo e muitos erros. Ressaltamos que o desempenho dos sujeitos em tal item não apresentou diferença significativa, segundo o teste *F* e de Tukey.

Dentro da categoria “leitura dos dados” tivemos apenas a Q.2.1., cujo número de itens errados supera o número de corretos, além de ser o único que não apresenta taxa de itens em branco. A taxa de desempenho na categoria “leitura dos dados” (77,98%), só não foi melhor, devido à quantidade de erros em tal item.

Na Figura 32, encontraremos um exemplo de tal item. Nele, solicitávamos aos sujeitos que observassem dois quadros com gráficos - um de colunas (a) e os outros de setor (b) - para que tomassem a decisão sobre qual deles leriam para saber a quantidade de turistas que se hospedou em hotéis no ano de 2001.

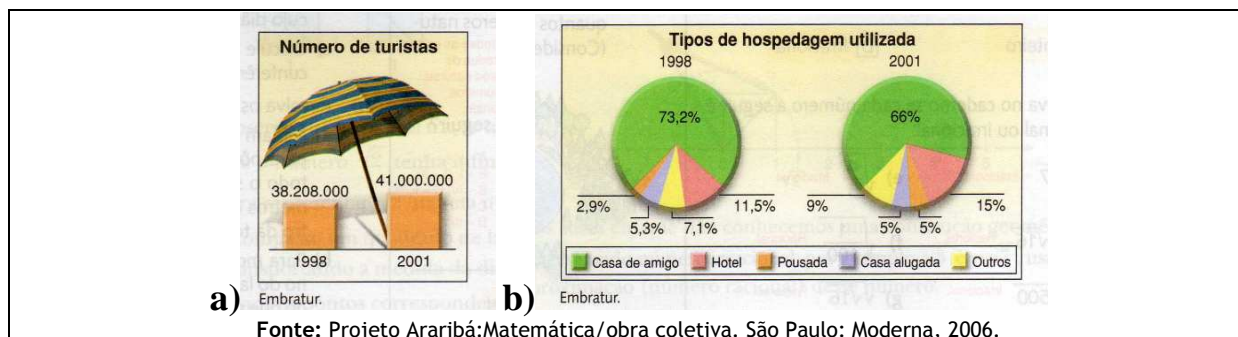


Figura 32: Exemplo dos gráficos utilizados na questão Q.2.1.

Nesse sentido, os sujeitos deveriam fazer a opção pelo gráfico (a), uma vez que este traz o título “número de turistas” enquanto o outro (b) refere-se aos “tipos de hospedagem utilizada”.

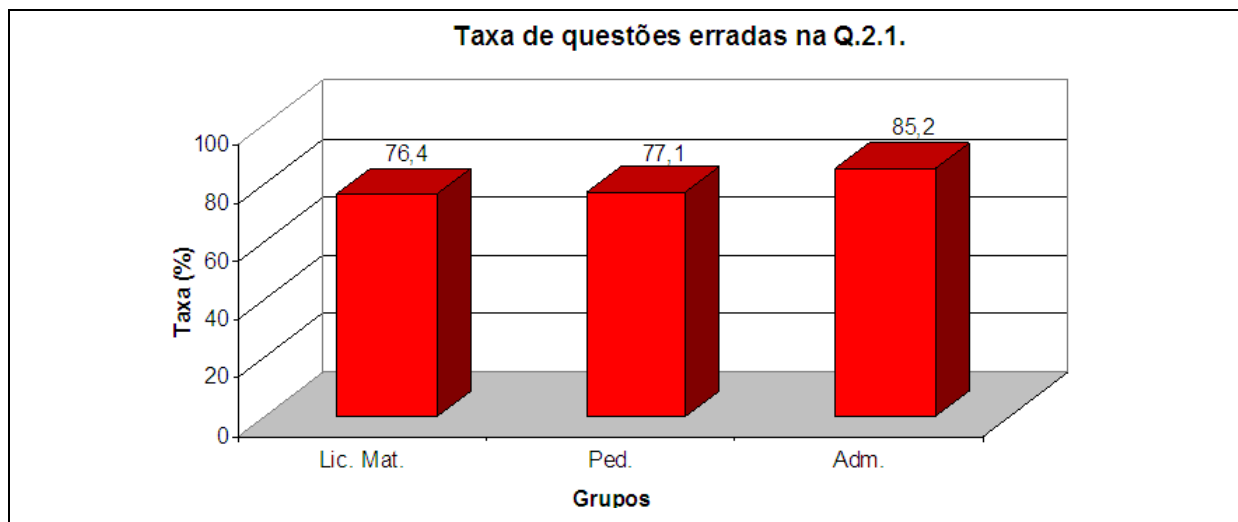


Figura 33: Gráfico de colunas com a Taxa de respostas erradas à questão Q.2.1 no nível “leitura dos dados”, segundo os grupos.

A partir do resultado total dos itens errados registrados no gráfico da Figura 33, podemos notar que tivemos 138 sujeitos que fizeram a opção por alguma das alternativas erradas e desses 138, 102 sujeitos (58,6%) responderam que era necessário ler os dois gráficos (a) e (b). Consideramos esse, um índice alto para um item que exigia somente uma leitura simples e pontual dos dados.

Uma possível explicação para tal resposta pode estar no fato de que os sujeitos não atentaram para os títulos dos gráficos. Tal argumento ganha força ao analisarmos os tipos de erros cometidos pelos sujeitos nas questões que solicitavam construções de gráficos. O que constatamos foi que o erro mais comum era a ausência de título.

Com vistas a obter maiores informações sobre o comportamento dos grupos, passaremos a discutir os seus desempenhos por itens, ainda considerando cada uma das classificações propostas por Curcio (leitura dos dados, leitura entre os dados e leitura além dos dados). Olharemos em particular as questões que apresentaram diferenças significativas, segundo o teste qui-quadrado (ver Tabela 7 a seguir).

Tabela 7: Desempenho dos sujeitos nas questões de Leitura e Interpretação de Gráficos no nível “leitura dos dados”, segundo os grupos e valor do teste qui-quadrado.

Itens	Grupos	Taxa de Acerto (%)			Teste qui-quadrado	
		Certo	Errado	Branco	$\chi^2(2)$	p-valor
Q.2.1	Lic. Mat.	23,6	76,4	0	1,655	0,437
	Pedag.	22,9	77,1	0		
	Adm.	14,8	85,2	0		
Q.2.2	Lic. Mat.	94,4	5,6	0	13,976	<b>0,001</b>
	Pedag.	70,8	25,0	4,2		
	Adm.	88,9	9,3	1,9		
Q.2.3	Lic. Mat.	90,3	8,3	1,4	4,201	0,122
	Pedag.	77,1	16,7	6,3		
	Adm.	87,0	13,0	0		
Q.3.a.a	Lic. Mat.	93,1	5,6	1,4	4,488	0,106
	Pedag.	87,5	0	12,5		
	Adm.	98,1	0	1,9		
Q.3.a.b	Lic. Mat.	79,2	19,4	1,4	49,541	<b>0,000</b>
	Pedag.	16,7	66,7	16,7		
	Adm.	68,5	29,6	1,9		
Q.3.c	Lic. Mat.	93,1	4,2	2,8	80,023	<b>0,000</b>
	Pedag.	64,6	10,4	25,0		
	Adm.	88,9	7,4	3,7		
Q.6.2	Lic. Mat.	72,2	13,9	13,9	6,687	<b>0,035</b>
	Pedag.	50,0	2,1	47,9		
	Adm.	68,5	22,2	9,3		

Valores em negrito indicam que os desempenhos foram estatisticamente significativos segundo o teste qui-quadrado.

O item Q.2.2 era de múltipla escolha e requeria que o sujeito lesse o gráfico (a) (vide figura 32) e dele retirasse as informações necessárias para selecionar qual das alternativas trazia a opção correta. Não estamos falando de uma leitura elaborada, mas sim da localização de uma informação dentro do registro gráfico. Observe a Tabela 7 que mostra a taxa de desempenho dos sujeitos:

A diferença percentual na taxa de acertos do GFP fica aquém da apresentada pelo GFLM e, principalmente pelo GFLM (18,1 e 23,6 pontos percentuais, respectivamente). Assim constatamos, também apoiados nos resultados obtidos com o teste de Tukey, que foi o desempenho do GFP o responsável pela diferença significativa entre os resultados dos grupos, já que não houve nível de significância estatística entre os desempenhos dos GFADM e GFLM.

No item 3.a.b, foi solicitado que os sujeitos fizessem a opção entre dois gráficos de barras empilhadas para responder se havia mais homens ou mulheres no Brasil, na faixa etária de 0 a 4 anos no ano de 1996.

Tratava-se de um item de baixo custo cognitivo, porém, mais uma vez, o GFP foi o responsável por haver diferença altamente significativa entre os grupos. De fato, nota-se uma diferença no percentual de acerto desse grupo para os outros dois bastante grande (51,8 pontos percentuais para o GFADM e 62,5 para o GFLM).

Para facilitar a lembrança, trazemos a seguir o item 3.a.b, acompanhado da resposta de um sujeito do GFP.

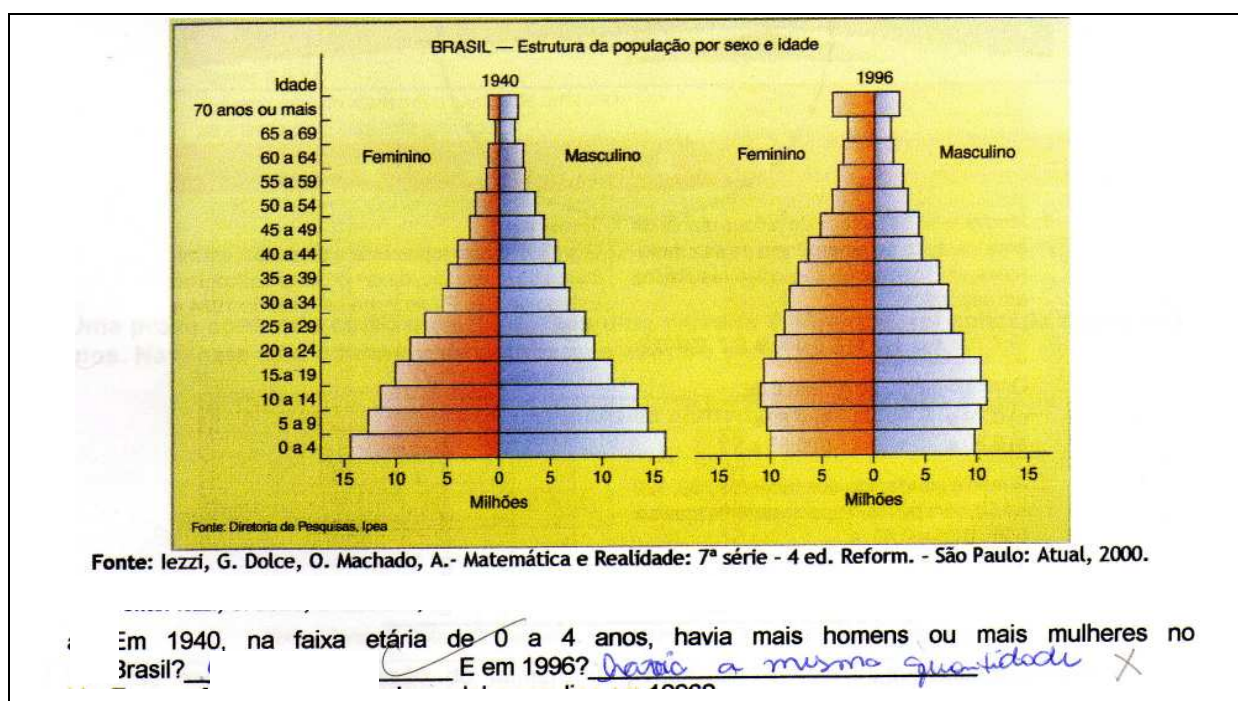


Figura 34: Exemplo de respostas dos sujeitos para o item Q.3.a.b.

Refletindo sobre o item, notamos que do ponto de vista perceptivo, a diferença entre homens e mulheres, na faixa de 0 a 4 anos de idade, era muito pequena, no entanto do ponto de vista numérico ela estava bem presente e podia ser extraída lendo-se o eixo "x" do gráfico da esquerda. Assim, conjecturamos que o grupo GFP tinha dificuldade em ler gráfico não convencional (no caso, barras empilhadas, lembrando uma pirâmide), mesmo que fosse para fazer uma leitura pontual dos dados. O exemplo apresentado na Figura 34 apoia nossa conjectura.

O item Q.3.c solicitava que o sujeito, com base no mesmo gráfico de barras empilhadas, apontasse qual a faixa etária que tinha menos população em 1940. Como todos os itens referentes a esse nível de leitura, este também não apresentava alto grau de dificuldade, uma vez que se tratava de uma localização de informação. Alias, acreditamos que este item é um pouco mais fácil que o anterior porque ele pode ser respondido apenas a partir da percepção – qual das barras empilhadas é a menor. E, de fato, todos os grupos se saíram bem (o GFLM chegou a apresentar 93,1 % de acerto, o GFADM 88,9% e o GFP apresentou 64,6%), embora, novamente, o GFP obtém percentual demarcadamente inferior aos outros dois grupos (28,5 pontos percentuais aquém do GFLM e 24,3 pontos percentuais de GFADM), o qual também foi estatisticamente significativo segundo o teste de Tukey.

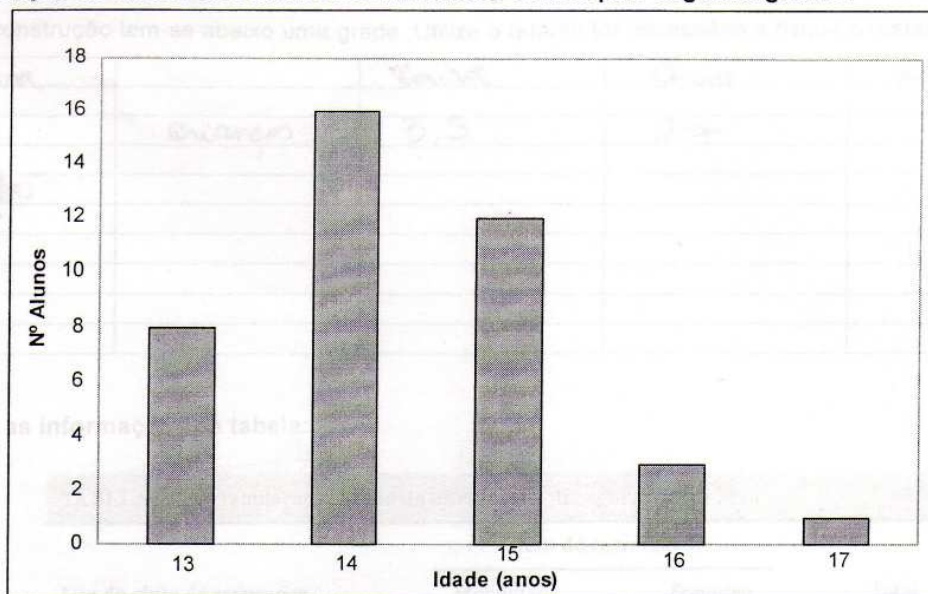
Salientamos que este item não exigia um alto custo cognitivo, e, por essa razão, não esperávamos uma taxa alta de questões em branco no grupo dos futuros Pedagogos.

O item Q.6.2 solicitava que os sujeitos, por meio de um gráfico de colunas representando a distribuição de idades dos alunos de uma classe, calculassem a idade modal. Essa questão está enquadrada no nível “leitura dos dados” por se tratar da localização de uma informação explícita no gráfico e de pouca exigência cognitiva. O GFP foi, mais uma vez, o responsável por haver diferença significativa entre os grupos, já que o teste de Tukey não aponta nível de significância estatística entre os desempenhos dos GFADM e GFLM.

De fato, encontramos os números de acertos do GFP distando do GFLM em 22,2 pontos percentuais e do GFADM em 18,5 pontos percentuais, diferenças significativas para uma questão considerada fácil, além de encontramos esse mesmo grupo de Pedagogos com um alto índice de questões em branco. Os desempenhos dos sujeitos diferiram significativamente entre os grupos [ $F(2, 135) = 2,536; p = 0,083$ ].

Apesar de não encontrarmos as estratégias registradas no corpo do teste, quando analisamos os itens errados do GFADM, percebemos que cerca de 50% das respostas erradas são com o valor “15”, como exemplificado na Figura 35.

6. A distribuição de idades dos alunos de uma classe é dada pelo seguinte gráfico:



Fonte: Iezzi, G. Dolce, O. Machado, A. - Matemática e Realidade. São Paulo:Atual, 2000.

6.2. A idade modal é:

15 anos X

Figura 35: Exemplo de respostas dos sujeitos ao item Q.6.2.

Acreditamos que os sujeitos tendem a fazer a leitura pontual buscando a “coluna do meio” como resposta para a Moda. Nesse caso, além do erro conceitual de Moda e de confundi-la com Mediana, também apresentam o erro conceitual de Mediana.

#### b) Categoria: Leitura entre os dados

Essa categoria contava com dois itens, Q.3.b e Q.6.3. Dessas, discutiremos os itens que apresentam uma diferença significativa, segundo o teste qui-quadrado.

Tabela 8: Desempenho dos sujeitos nos itens de Leitura e Interpretação de Gráficos no nível “leitura entre os dados”, segundo os grupos e valor do teste qui-quadrado

Interpretação. de Gráficos (Nível Leitura entre os	Questões	Grupos	Taxa de Acerto (%)			Teste qui-quadrado	
			Certo	Errado	Branco	$\chi^2(2)$	p-valor
			<b>Q.3.b</b>	Lic. Mat.	70,8	15,3	13,9
Pedag.	43,8	10,4		45,8			
Adm.	59,3	29,6		11,1			
<b>Q.6.3</b>	Lic. Mat.	73,6	52,1	81,5	102,519	<b>0,000</b>	
	Pedag.	6,9	0,0	3,7			
	Adm.	19,4	47,9	14,8			

Valores em negrito indicam que os desempenhos foram estatisticamente significativos segundo o teste qui-quadrado.

O item Q.3.b solicitava aos sujeitos que, em primeiro lugar, localizassem o gráfico de barras empilhadas referente ao ano de 1996, em segundo optassem pelo lado rosa (feminino) ou azul (masculino) e em terceiro, encontrassem a faixa que contivesse a idade modal. A questão em jogo era: em que faixa estava à idade modal masculina em 1996?

Trata-se de um item que exigia além da interpretação dos dados, a combinação e a integração das informações contidas no gráfico; por esse motivo enquadrado no nível “leitura entre os dados”. Como já dito na análise dos itens anteriores, o GFP é o responsável pela diferença significativa entre os grupos. Tal fato é comprovado por meio dos dados registrados na Tabela 8, onde encontramos a taxa de questões corretas do GFP distando do GFLM em 27 pontos percentuais e do GFADM em 15,5 pontos percentuais. Também detectamos, segundo o teste  $F$  que houve diferença significativa nos desempenhos entre os grupos na questão Q.3.b [ $F(2, 173) = 4,551; p = 0,012$ ].

Até este momento, pudemos comprovar que, quando analisamos os resultados com vistas nas taxas de questões em branco, as taxas do GFP são sempre as maiores quando comparadas as taxas dos outros dois grupos. Supomos então que, provavelmente, estes sujeitos desenvolveram uma atitude negativa em relação à Estatística, e, por esse motivo nem tentaram resolver as questões propostas ou não dominaram seus conteúdos.

Encontramos no exemplo da Figura 36 um sujeito do grupo dos futuros Pedagogos engrossando a taxa de questões em branco por não conhecer a palavra “modal”. Esse exemplo apoia nossa suposição.

Brasil? *homens / F em 1996?*  
 b) Em que faixa estava a idade modal masculina em 1996? *Não sei o que é modal.*

Figura 36: Exemplo de resposta de um sujeito do GFP ao item Q.3.b.

Quanto às taxas de itens errados, procuramos analisar algumas das respostas, na tentativa de encontrar as estratégias utilizadas. São elas:

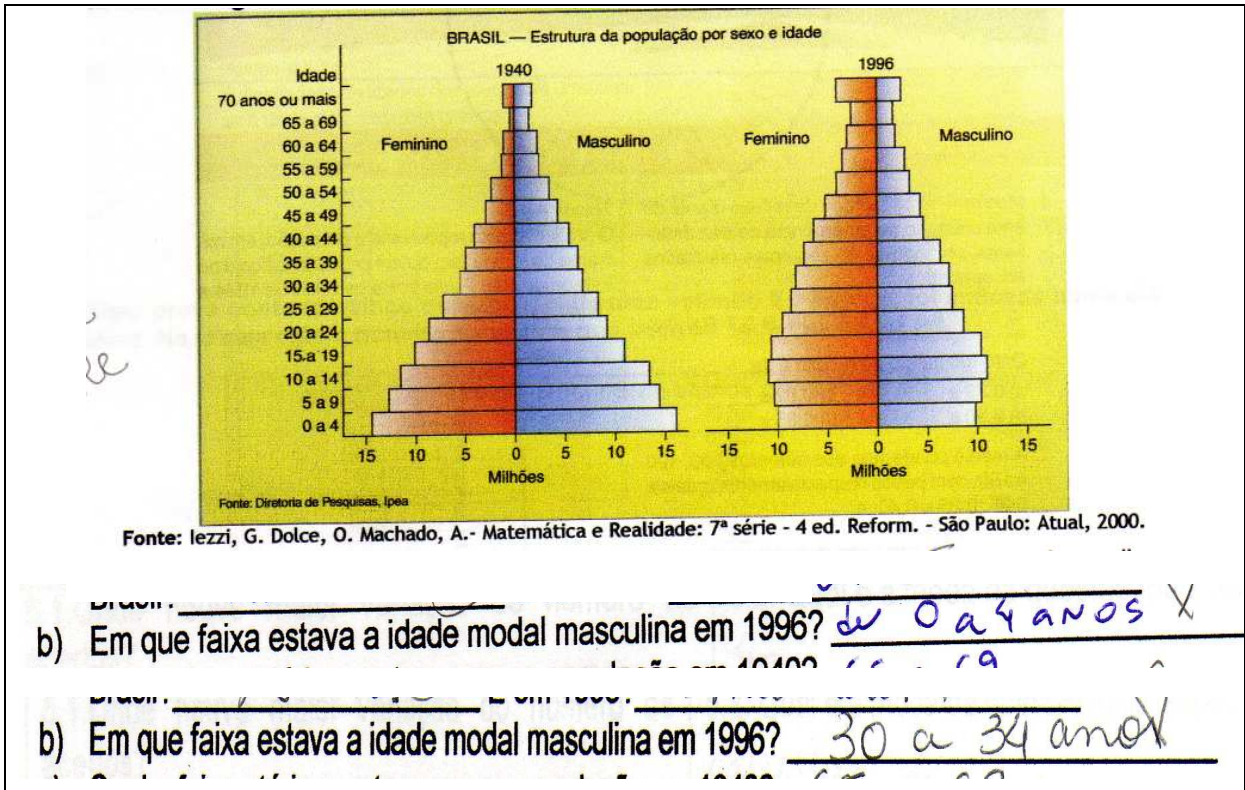


Figura 37: Exemplo de resposta dos sujeitos do GFADM ao item Q.3.b.

Para a resposta errada “de 0 a 4 anos” percebemos que os sujeitos não atentaram para o gráfico de barras da direita. Ele é que representa a estrutura da população por sexo e idade no ano de 1996. Supomos então, que estes sujeitos procuraram as barras maiores e as encontraram na base do gráfico empilhado do ano de 1940, que representa a faixa de “0 a 4” anos.

Para a resposta errada “de 30 a 34 anos”, acreditamos que os sujeitos “contaram” as faixas que representam as escalas das idades de “0 a 4” até “65 a 69”, perfazendo um total de 14 faixas. A faixa “30 a 34” anos ocupa o 7º lugar. Logo, levantamos a hipótese de que os sujeitos confundem a moda com mediana, além de cometerem erro conceitual tanto na Moda como na Mediana.

Já o item Q.6.3 solicitava que, a partir de um gráfico de colunas que representava a distribuição de idades dos alunos de uma classe, os sujeitos indicassem qual a idade que representava a mediana. Não houve diferença significativa nos desempenhos entre os três grupos na questão Q.6.3, segundo o teste de Tukey e teste  $F[F(2, 128)= 1,343; p=0,265]$ .

A diferença percentual na taxa de acertos dos três grupos fica aquém do esperado (tais taxas não passam de 7%), apesar de quando elaboramos a análise a priori, já esperávamos uma taxa baixa por causa do nível cognitivo exigido, no entanto contávamos com um índice maior. Também comprovamos que além da alta taxa de itens errados em cada grupo, também temos uma boa parcela de sujeitos que a deixou em branco; por exemplo, o grupo de futuros Pedagogos chega a quase 50%.

Apesar dos sujeitos não registrarem suas estratégias, acreditamos que aqueles que deram a resposta “13 anos”, raciocinavam pela altura das colunas e calculavam a mediana pela frequência, anotando a respectiva idade daquela coluna; já os que responderam “15 anos”, mostram uma tendência a procurar pela coluna localizada no meio do gráfico. Ambas as estratégias erradas nos remetem ao nível “leitura dos dados”, pois é o nível em que o indivíduo faz a leitura pontual.

Para percebermos tais estratégias erradas, trazemos na Figura 38, um exemplo de respostas de dois sujeitos ao item Q.6.3.

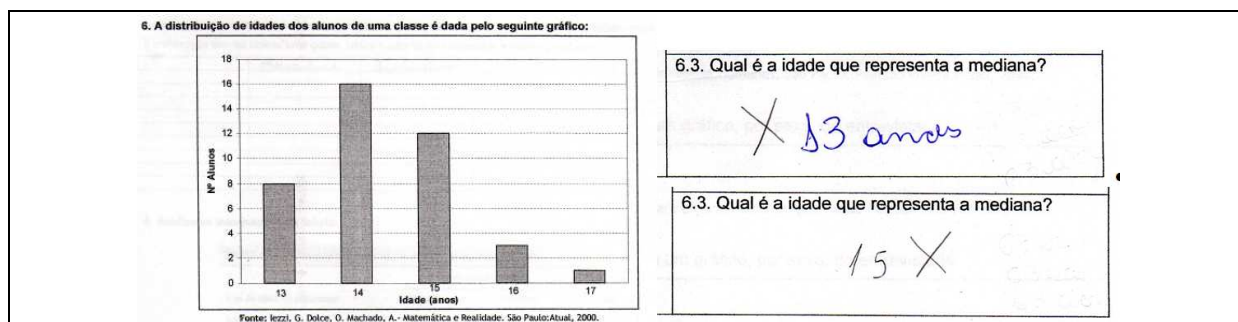


Figura 38: Exemplos de respostas dos sujeitos do GFLM e GFADM ao item Q.6.3.

### c) Categoria: Leitura além dos dados

Já o nível “leitura além dos dados” continha um único item - Q.6.1. Ele solicitava aos sujeitos que indicassem qual a idade média dos alunos a partir da leitura de um gráfico de colunas. Os desempenhos dos sujeitos diferiram

significativamente entre os grupos, tanto no teste  $F$  (ANOVA) como no teste qui-quadrado:  $[F(2, 138) = 6,346; p = 0,002]$ ,  $[\chi^2(1, N=139) = 85,475; p = 0,000]$ .

Apoiados nos resultados obtidos com o teste de Tukey, pudemos constatar que os desempenhos do GFP e GFADM foram os responsáveis pela diferença significativa entre os resultados dos grupos, apesar de todos os grupos demonstrarem um desempenho sofrível.

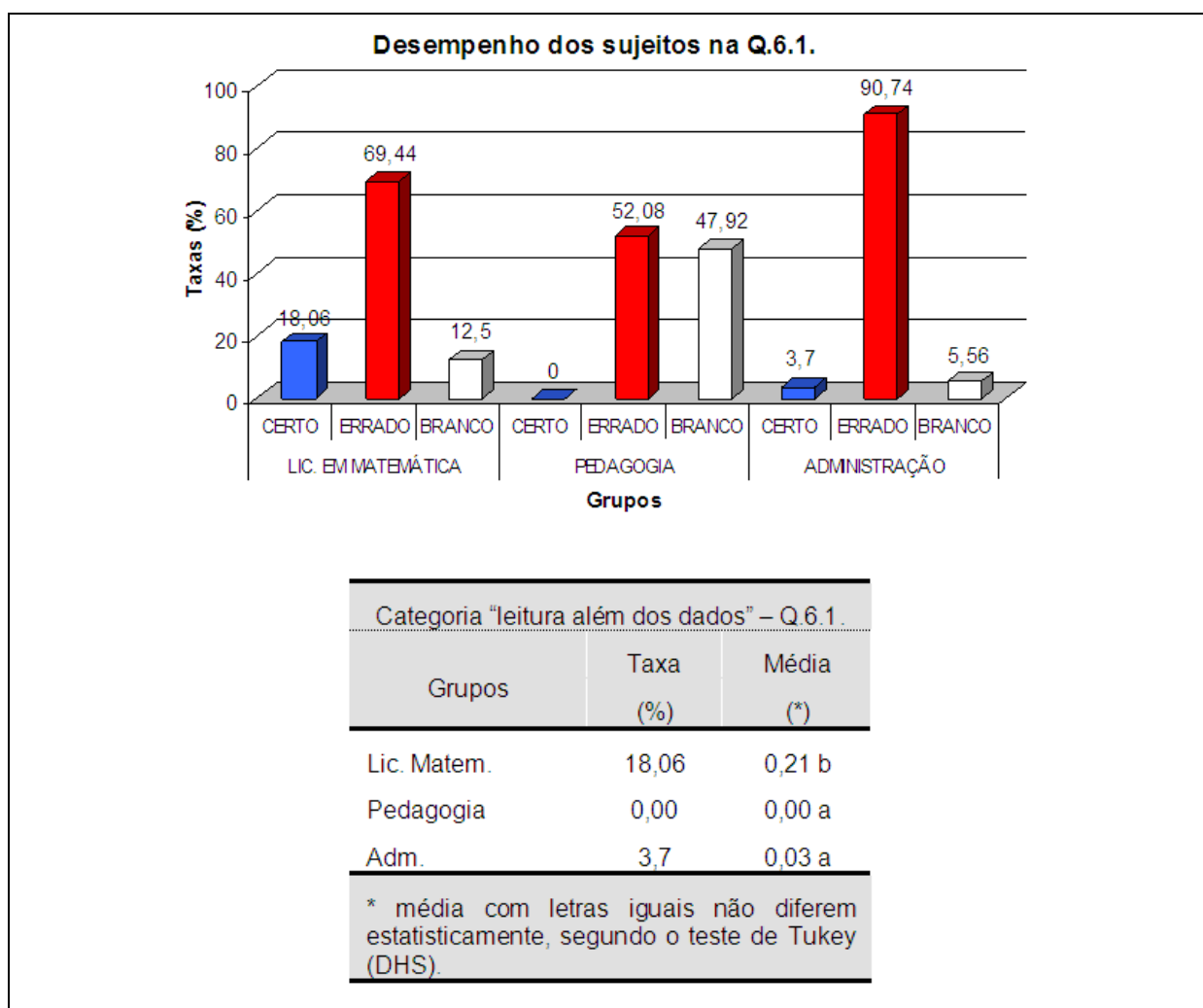


Figura 39: Desempenho dos sujeitos, segundo os grupos, no item Q.6.1 e tabela do teste de Tukey.

Os sujeitos do grupo de Pedagogia compõem o grupo que mais deixa esse item em branco, ao contrário de Licenciatura em Matemática e Administração, que detêm uma maior taxa de questões erradas, possivelmente na tentativa de acertar.

Ao observarmos as taxas de itens errados, acreditamos que a maior dificuldade apresentada ao resolver o problema foi justamente identificar que se

tratava de um gráfico de distribuição de idades dos alunos fictícios de uma classe. Os sujeitos não refletiram e não fizeram a interpretação de que existiam 8 alunos com 13 anos, 16 alunos com 14 anos, 12 com 15 anos, 3 com 16 e 1 aluno com 17 anos.

Por fim, na estratégia registrada por um sujeito do grupo de Licenciatura em Matemática na Figura 40, percebemos que este não atentou para o fato de ser uma média aritmética de dados agrupados.

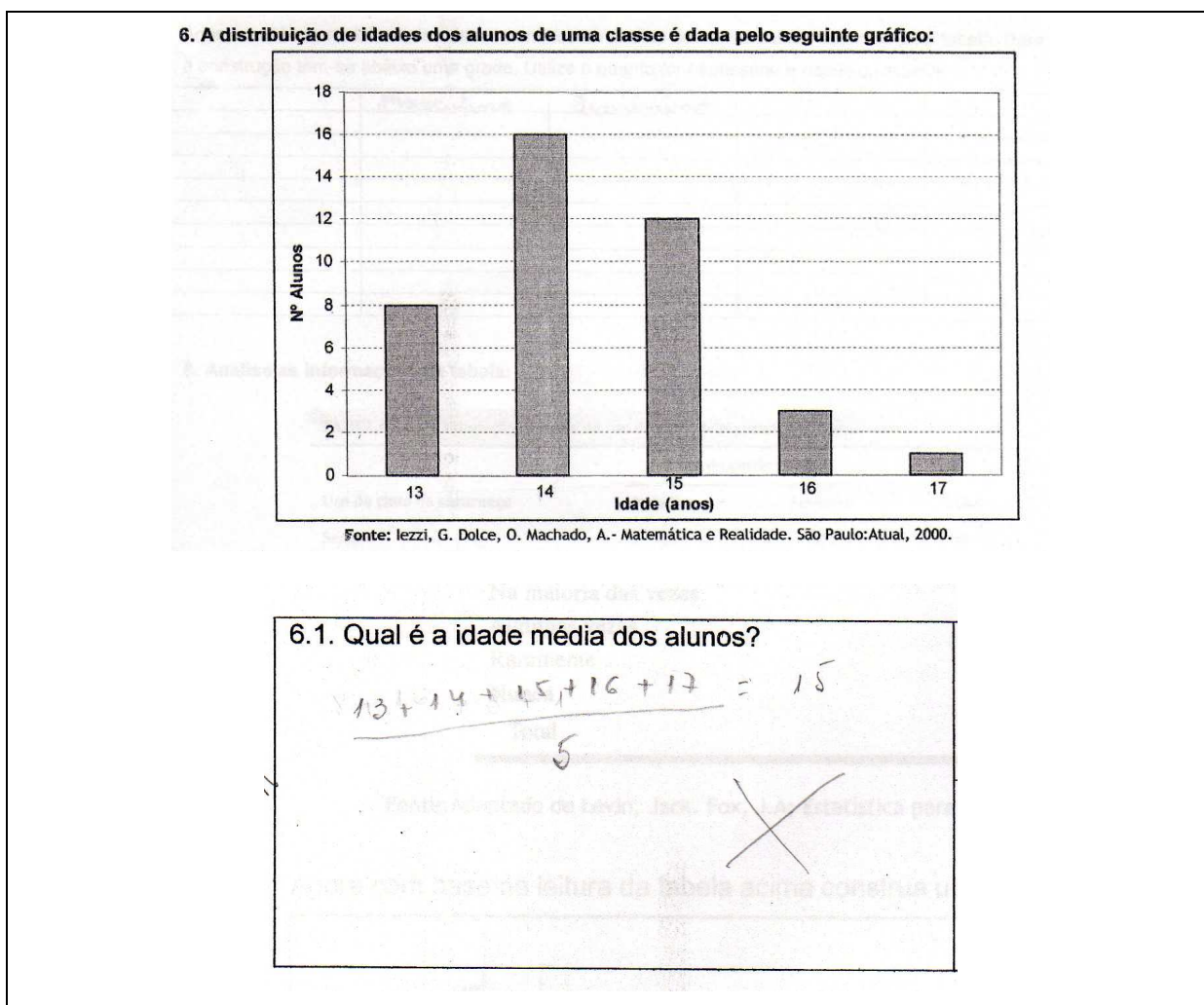


Figura 40: Exemplo de resposta de um sujeito do GFLM para o item Q.6.1.

Como essa resposta (15 anos) foi recorrente nos três grupos, supomos que estes sujeitos ainda não dominam a leitura de dados agrupados representados por gráficos.

#### 4.1.2.2. LEITURA E INTERPRETAÇÃO DE TABELAS E SUAS CATEGORIAS

Para essa unidade, o teste contava com 06 itens divididos da seguinte forma: 1 item aborda a categoria “nível básico”, 3 abordam o “nível intermediário” e 2 o “nível avançado”.

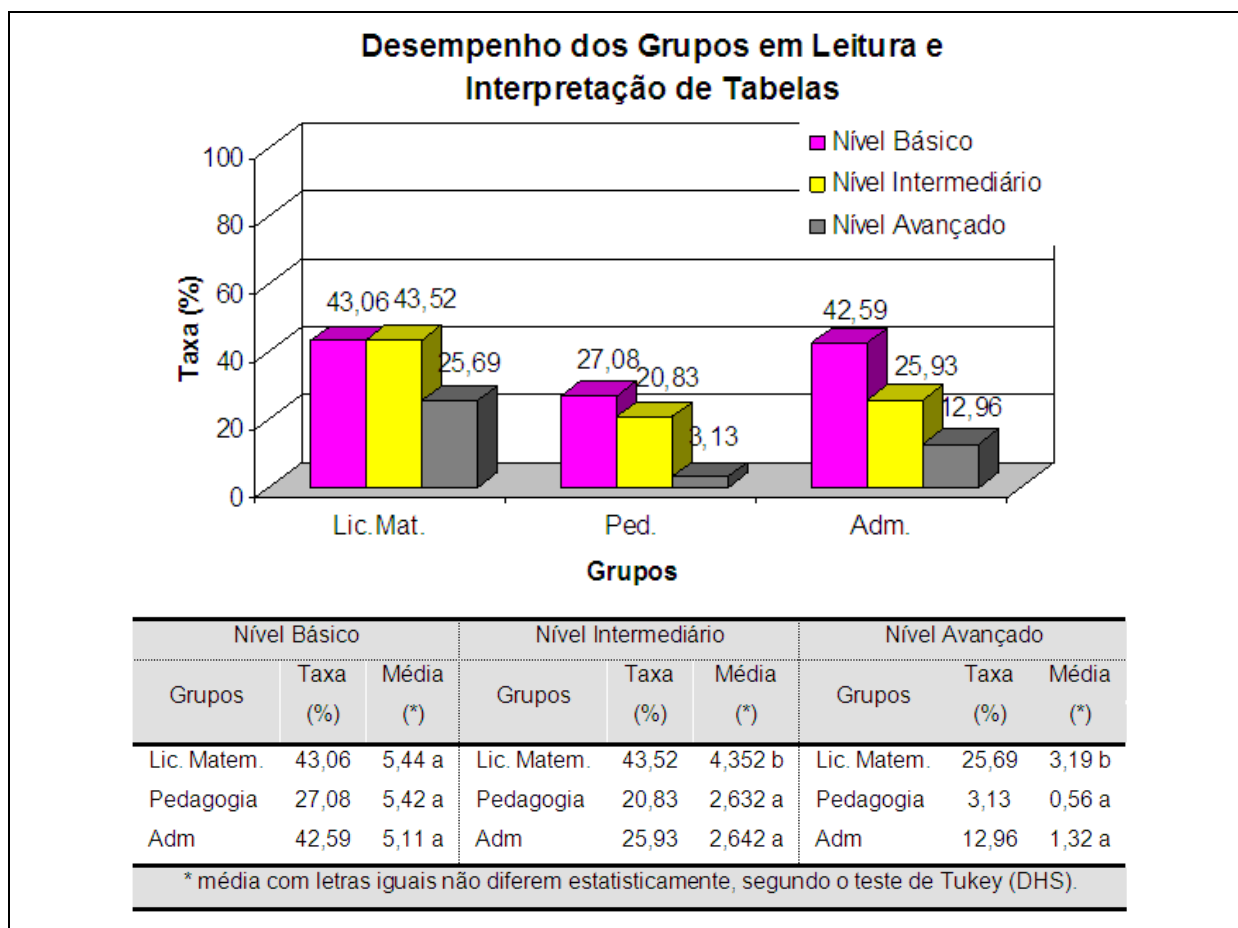


Figura 41: Desempenho dos sujeitos, segundo os grupos, nas categorias de Leitura e Interpretação de Tabelas e tabela com respectivos testes de Tukey.

Os desempenhos dos sujeitos nos itens de “nível básico” não diferiram significativamente entre os grupos [ $F(2, 123) = 0,059$ ;  $p = 0,943$ ]. Para os itens que abordavam o “nível intermediário” e “nível avançado”, respectivamente, os desempenhos dos sujeitos diferiram significativamente entre os grupos [ $F(2, 160) = 10,593$ ;  $p = 0,000$ ] e [ $F(2, 135) = 10,364$ ;  $p = 0,000$ ].

De acordo com o gráfico da Figura 41, os sujeitos foram bem mais sucedidos nos itens no “nível básico”, seguido do “nível intermediário” e do “nível avançado”. Podemos observar que, conforme o nível se torna mais complexo, a taxa de acertos tende a diminuir.

Os resultados apurados confirmam o que se espera, segundo a teoria de Wainer (1992). Para a *Leitura e Interpretação de Tabelas* encontramos as questões que somente extraem da tabela os dados que estão explícitos no “nível básico” e, portanto, com um bom índice de acerto por serem mais fáceis; já nas questões intermediárias que exigiam a interpolação ou a percepção da relação existente entre os dados, o desempenho foi menos positivo que o anterior; e, para o “nível avançado” que exigia um maior entendimento das estruturas dos dados, temos um desempenho menos positivo ainda.

Vamos nos deter na análise quantitativa dos itens arrolados nessa unidade, começando pelo “nível básico” e seu único item, Q.5.2., logo após o “nível intermediário” e “nível avançado”.

Como essa categoria trabalha com a representação tabular, achamos adequado apresentar todas as análises na forma de tabelas.

#### **a) Categoria: Nível Básico**

Essa categoria contava com apenas um item, Q.5.2, que trazia uma tabela simples com os números das questões de uma prova e o número de alunos que as acertaram. A tarefa solicitada era saber qual a Moda da questão mais respondida. Este item exigia somente que os sujeitos extraíssem da tabela os dados explícitos.

O teste qui-quadrado indicou que houve diferença significativa nos desempenhos dos sujeitos no item Q.5.2 [ $\chi^2 (1, N=174) = 9,195; p = 0,002$ ]. Já, segundo o teste *F*, os desempenhos dos sujeitos não diferiram entre os grupos [ $F(2, 173) = 1,835; p = 0,163$ ].

Observe a Tabela 9 que apresenta os resultados atingidos nesse item.

Tabela 9: Desempenho dos sujeitos, segundo os grupos, no item Q.5.2 no “Nível Básico”.

QUESTÃO Q.5.2.	GRUPOS						TOTAL	
	LIC. MAT		PEDAG.		ADM		N	%
	N	%	N	%	N	%		
CERTO	31	43,1	13	27,1	23	42,6	67	38,5
ERRADO	26	36,1	11	22,9	22	40,7	59	33,9
BRANCO	15	20,8	24	50,0	9	16,7	48	27,6

Quando examinamos as taxas de acertos em cada grupo, verificamos que estas, não passam de 50% em nenhum dos grupos, além de serem seguidas de perto pela taxa de itens erradas.

Apontamos como provável causa dos 33,9% de itens errados, o fato dos sujeitos não atentarem para qual coluna da tabela trazia a variável independente. Parece que estes mesmos sujeitos não refletiram sobre o questionamento do enunciado do problema - qual é a Moda da questão mais respondida. No lugar de responderem a variável independente “questão 2”, eles responderam a variável dependente “40”. Observem na Figura 42, dois exemplos de respostas dos sujeitos para esse item.

**5. Uma prova contendo cinco questões, cada uma valendo 2,0 pontos, foi aplicada numa classe de 40 alunos. Na tabela está o número de alunos que acertou cada questão:**

Questão	Número de alunos que acertaram
1	12
2	40
3	24
4	32
5	36

Fonte: lezzi, G. Dolce, O. Machado, A.- Matemática e Realidade. São Paulo:Atual, 2000.

5.2. Qual é a moda da questão mais respondida?

40

5.2. Qual é a moda da questão mais respondida?

A moda é acertar.

Figura 42: Exemplos de respostas dos sujeitos ao item Q.5.2.

## b) Categoria: Nível Intermediário

O nível intermediário contava com três itens: Q.4.1.a, Q.4.1.b e Q.4.2. Discutiremos os itens que apresentam uma diferença significativa, segundo o teste qui-quadrado, individualmente.

Tabela 10: Desempenho dos sujeitos, segundo os grupos, nas questões no “Nível Intermediário

Leitura e Interpretação, de Tabelas (Nível Intermediário)	Questões	Grupos	Taxa de Acerto (%)			Teste qui-quadrado	
			Certo	Errado	Branco	$\chi^2(1)$	p-valor
Q.4.1.a	Lic. Mat.	31,9	54,2	13,9	53,289	<b>0,000</b>	
	Pedag.	4,2	72,9	22,9			
	Adm.	11,1	87,0	1,9			
Q.4.1.b	Lic. Mat.	16,7	43,1	40,3	72,681	<b>0,000</b>	
	Pedag.	0	52,1	47,9			
	Adm.	1,9	92,6	5,6			
Q.4.2	Lic. Mat.	81,9	6,9	11,1	79,317	<b>0,000</b>	
	Pedag.	58,3	4,2	37,5			
	Adm.	64,8	18,5	16,7			

Valores em negrito indicam que os desempenhos foram estatisticamente significativos segundo o teste qui-quadrado.

No item Q.4.1.a era solicitado aos sujeitos que a partir de uma tabela simples, eles respondessem se era possível ou não calcular a média daqueles dados, assinalando a alternativa com a palavra “sim” ou com a palavra “não”. A tabela representava as opiniões de 60 pessoas a respeito de um filme que acabava de estreiar numa cidade fictícia.

Os desempenhos dos sujeitos diferiram significativamente entre os grupos no item Q.4.1.a [ $F(2, 151) = 10,299; p = 0,000$ ]. O teste de Tukey (DHS) indicou que os desempenhos dos sujeitos de Pedagogia e de Administração foram significativamente menos positivos que aquelas apresentadas pelos sujeitos de Licenciatura em Matemática. De fato, apoiados no teste de Tukey, nota-se que a diferença percentual na taxa de acertos do GFP e GFADM fica aquém da apresentada pelo GFLM (27,7 e 20,8 pontos percentuais respectivamente). Logo, esses dois grupos foram os responsáveis pela diferença significativa entre os resultados dos grupos.

Observe a Tabela 11 e as taxas de desempenho dos sujeitos.

Tabela 11: Desempenho dos sujeitos, segundo os grupos, no item Q.4.1.a

QUESTÃO Q.4.1.a.	GRUPOS						TOTAL	
	LIC. MAT		PEDAG.		ADM		N	%
	N	%	N	%	N	%		
CERTO	23	31,9	2	4,2	6	11,1	31	17,8
ERRADO	39	54,2	35	72,9	47	87,0	121	69,5
BRANCO	10	13,9	11	22,9	1	1,9	22	12,6

Esse item exigia uma percepção da relação existente entre as variáveis nominais e o tipo de cálculo de medida central que esta aceita, ou seja, somente a Moda. Notamos pelos desempenhos dos grupos que essa relação não foi percebida.

Novamente chamamos a atenção para o índice de questões em branco do GFP e para o fato de que o GFADM apresenta o maior índice de questões erradas, porém o menor de questões em branco. Então, podemos dizer que o GFADM erra na tentativa de acertar. Já a taxa de acerto do GFLM ficou aquém dos 50%; esperávamos uma taxa de acerto mais positiva, uma vez que estes, a priori, possuem um arcabouço matemático superior aos demais.

Para esse item, podemos levantar a hipótese que 69,5% dos sujeitos pesquisados ainda não compreendem as relações das variáveis qualitativas com a medida de tendência central “moda”.

Já o item Q.4.1.b solicitava aos alunos que explicassem por extenso o porquê é possível ou não calcular a média dos dados da tabela (variáveis nominais).

Tabela 12: Desempenho dos sujeitos, segundo os grupos, no item Q.4.1.b.

QUESTÃO Q.4.1.b.	GRUPOS						TOTAL	
	LIC. MAT		PEDAG.		ADM		N	%
	N	%	N	%	N	%		
CERTO	12	16,7	0	0	1	1,9	13	7,5
ERRADO	31	43,1	25	52,1	50	92,6	106	60,9
BRANCO	29	40,3	23	47,9	3	5,6	55	31,6

O sucesso do item Q.4.1.b estava atrelado à questão anterior e por esse motivo, a taxa de itens errados foi menos positivo ainda. Observamos que o índice de itens em branco aumentou porque muitos sujeitos que assinalaram a Q.4.1.a, deixavam a explicação em branco. Já, nos itens corretos, os índices diminuíram pelo motivo dos sujeitos assinalarem a opção correta, porém, explicarem sua escolha de uma forma incorreta.

Os desempenhos dos sujeitos diferiram significativamente entre os grupos no item Q.4.1.b [ $F(2, 118) = 11,732; p = 0,000$ ]. O teste de Tukey (DHS) indicou que os desempenhos dos sujeitos de Pedagogia e de Administração foram significativamente menos positivos que aquelas apresentadas pelos sujeitos de Licenciatura em Matemática. Novamente temos os grupos GFP e GADM sendo os responsáveis pela diferença significativa entre os resultados dos grupos. Os desempenhos do três grupos ficam aquém de serem considerados medianos, no entanto, o GFADM e principalmente o GFP obtiveram um desempenho, no mínimo, sofrível.

Na tentativa de buscar explicação para tal taxa de itens corretos, buscamos no protocolos os tipos de erros cometidos pelos sujeitos. Observe a Figura 43 que traz alguns exemplos de respostas corretas “não é possível calcular”; no entanto, a explicação está errada.

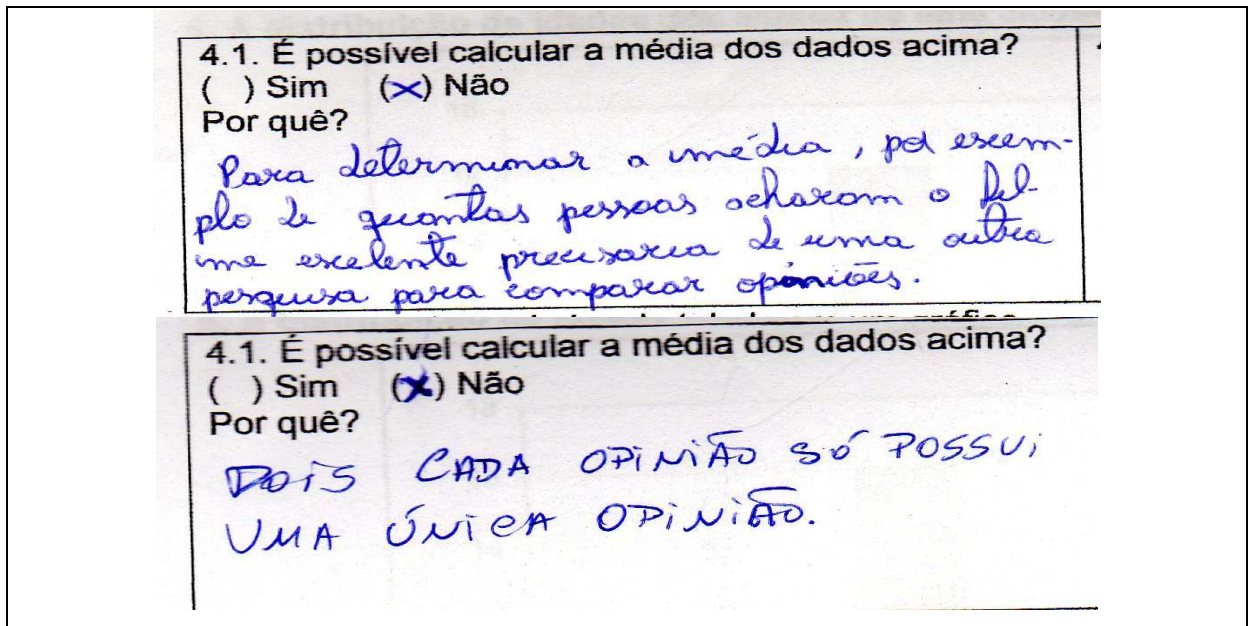


Figura 43: Exemplos de respostas dos sujeitos no item Q.4.1.a e Q.4.1.b.

Agora, observe a Figura 44 que traz dois exemplos de respostas: uma resposta correta com a explicação “em branco” e, outra resposta totalmente errada.

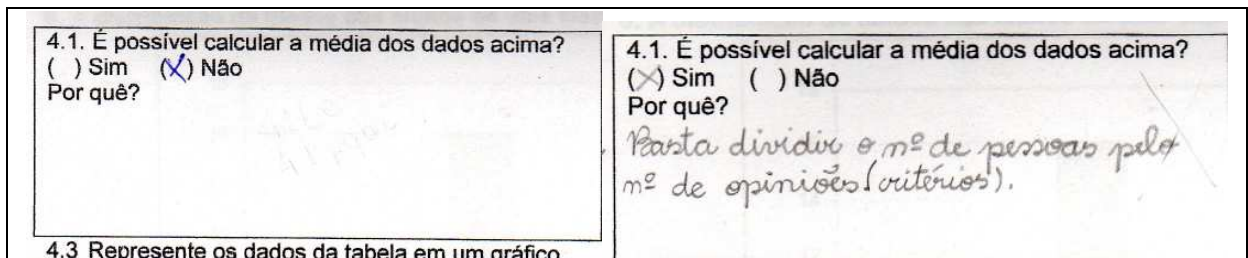


Figura 44: Exemplos de respostas dos sujeitos ao item Q.4.1.a e Q.4.1.b

O item Q.4.2 solicitava que os sujeitos, baseado na mesma tabela simples, indicassem a opinião mais frequente (Moda). Esse item foi classificado no nível intermediário por exigir uma percepção da relação do cálculo da Moda com os dados da tabela. O fato é que para solucionar o problema, é necessário conhecer o conceito de Moda.

Os desempenhos dos sujeitos diferiram significativamente entre os grupos no item Q.4.2 [ $F(2, 138) = 3,182; p = 0,045$ ]. O teste de Tukey (DHS) indicou que não houve diferença entre os desempenhos dos grupos.

Tabela 13: Desempenho dos grupos no item Q.4.2 no “Nível Intermediário”.

QUESTÃO Q.4.2	GRUPOS						TOTAL	
	LIC. MAT		PEDAG.		ADM			
	N	%	N	%	N	%	N	%
CERTO	59	81,9	28	58,3	35	64,8	122	70,1
ERRADO	5	6,9	2	4,2	10	18,5	17	9,8
BRANCO	8	11,1	18	37,5	9	16,7	35	20,1

Pelos dados na Tabela 13, podemos observar que 70,1% dos sujeitos foram bem sucedidos nessa questão. Não se tratava de um item de nível cognitivo muito alto, porém, necessitava que os sujeitos compreendessem a relação existente entre a medida de tendência central “Moda” e a opinião mais computada “bom”.

Mesmo assim, temos cerca de 30% dos sujeitos que erram ou deixam o item em branco. Nos exemplos da Figura 45, podemos observar duas respostas erradas.

OPINIÃO	Nº DE PESSOAS
Excelente	9
Ótimo	15
Bom	18
Regular	12
Ruim	3
Péssimo	3
<b>TOTAL</b>	<b>60</b>

Fonte: Iezzi, G. Dolce, O. Machado, A.- Matemática e Realidade: 6ª série - São Paulo:Atual, 2000.

4.2. Indique a opinião mais freqüente (moda).

*Ruim e péssimo*

4.2. Indique a opinião mais freqüente (moda).

*3*

Figura 45: Exemplos de respostas dos sujeitos ao item Q.4.2.

Analisando esses dois tipos de erros, levantamos a hipótese de que como conceitualmente a “moda” é o valor mais frequente, os sujeitos olharam quais os números que se repetiam na coluna “nº de pessoas”, encontraram o número “3” que

repete duas vezes e responderam o próprio número ou as respectivas opiniões “ruim e péssimo”.

Nesse sentido, apontamos para um erro conceitual na medida de tendência central - Moda – e para a falta da percepção de qual era a variável em jogo: a opinião ou o número de pessoas?

### c) Categoria: Nível Avançado

Essa categoria trazia dois itens, Q.5.1 e Q.9.e. Discutiremos os itens que apresentam uma diferença significativa, segundo o teste qui-quadrado, individualmente.

Tabela 14: Desempenho dos sujeitos, segundo os grupos, nas questões no “Nível Avançado”.

Leitura e Interpretação de Tabelas (Nível Avançado)	Questões	Grupos	Taxa de Acerto (%)			Teste qui-quadrado	
			Certo	Errado	Branco	$\chi^2(1)$	p-valor
			Q.5.1	Lic. Mat.	18,1	52,8	29,2
Pedag.	2,1	47,9		50,0			
Adm.	5,6	79,6		14,8			
Q.9	Lic. Mat.	33,3	23,6	43,1	57,471	<b>0,000</b>	
	Pedag.	4,2	43,8	52,1			
	Adm.	20,4	70,4	9,3			

Valores em negrito indicam que os desempenhos foram estatisticamente significativos segundo o teste qui-quadrado.

Os desempenhos dos sujeitos diferiram significativamente entre os grupos no item Q.5.1 [ $F(2, 120) = 5,100; p = 0,008$ ]. O teste de Tukey (DHS) indicou que os desempenhos dos sujeitos de Pedagogia e Administração foram significativamente menos positivos que aqueles apresentados pelos sujeitos de Licenciatura em Matemática, conseqüentemente os dois grupos que foram responsáveis por tal diferença.

No item Q.5.1, era solicitado aos sujeitos que, baseados em uma tabela de dados agrupados, apontassem onde houve maior variação do número de acertos.

Apesar de ser necessária uma percepção da relação existente entre os dados desse item, fato este que o caracterizaria no nível intermediário, ele aborda um maior entendimento da estrutura dos dados (a variável dependente e a

independente) e uma visão global da tabela portanto enquadrada como nível intermediário.

Acreditamos que a taxa de questões erradas foi elevada por causa da falta de discernimento entre a indicação de onde estava a variação da questão nº 1 para a questão nº 2. Nesses casos, consideramos erradas aquelas respostas em que não eram indicados de que questão para qual questão que havia a variação; ou, indicavam da variável independente para a dependente.

O exemplo da Figura 46 apoia nossa conjectura.

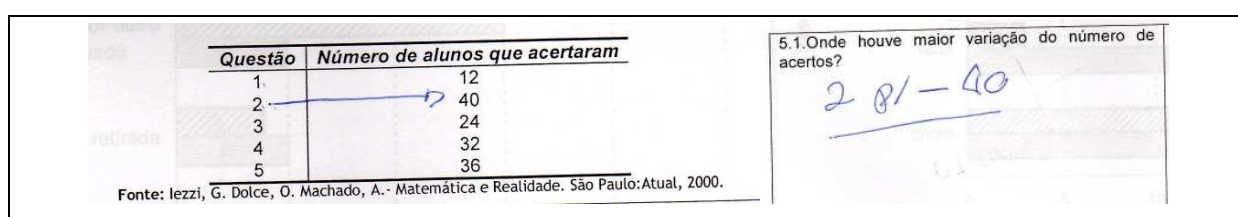


Figura 46: Exemplo de resposta de um sujeito do GFLM ao item Q.5.1.

No item Q.9 os desempenhos dos sujeitos diferiram significativamente entre os grupos [ $F(2, 112) = 12,396; p = 0,000$ ]. O teste de Tukey (DHS) indicou que os desempenhos dos sujeitos de Pedagogia e Administração foram significativamente menos positivos que aqueles apresentados pelos sujeitos de Licenciatura em Matemática, logo os dois grupos foram os responsáveis por tal diferença significativa entre os resultados dos grupos.

O enunciado desse item contava com uma tabela de dupla entrada estratificada com três categorias – gênero, região e grau de instrução. Tal tabela foi extraída do teste do ENEM do ano de 2001. Trata-se de uma leitura no nível avançado, porque o sujeito analisaria questões implícitas como, a quantidade de anos que um indivíduo estudou com o grau de instrução completo ou incompleto e ainda as regiões geográficas dos estados brasileiros em questão. Também privilegiava a visão global da tabela, uma vez que precisariam cruzar todas as colunas com as linhas para obter a alternativa correta.

Como já era esperado na análise a priori do teste, esse item teria um baixo desempenho, pelo motivo de envolver muitas variáveis cognitivas, além de ser uma questão que não contava só com os cálculos matemáticos e estatísticos.

Para essa questão, levantamos a hipótese de que as tabelas de dupla entrada e principalmente aquelas que envolvam outras informações, além das matemáticas, estão sendo negligenciadas no rol de conteúdos estatísticos ensinados nas instituições; haja vista que a taxa de questões erradas é praticamente o dobro da taxa de questões corretas, sem contar que a taxa de questões em branco ultrapassam a de questões corretas.

Por ser uma questão de múltipla escolha, não será possível analisar os tipos de erros.

#### **4.1.2.3. REPRESENTAÇÃO SEMIÓTICA EM ESTATÍSTICA E SUAS CATEGORIAS**

Essa categoria trabalhava com a conversão e o tratamento baseados na teoria Registro de Representação Semiótica, discutida no Capítulo II. Resgatando as idéias de Duval (2005), a conversão é a transformação de um registro de representação em outro registro, conservando a totalidade ou parte do objeto matemático trabalhado; e o tratamento é a transformação de uma representação em outra representação, porém ficando no mesmo registro.

Em seu artigo, Flores e Moretti (2006) ressaltam que a mudança de um registro em outro não é tão simples como possa parecer e que o ensino da Estatística privilegia muito mais a leitura e identificação de dados registrados em tabelas ou gráficos do que a própria construção destas representações com todos seus elementos.

Nesse sentido, nossa análise dos desempenhos dos sujeitos nessa categoria levará em conta, principalmente, a construção correta de todos os elementos de tabelas e gráficos no momento da conversão ou do tratamento dos registros.

Para a construção de tabelas, consideraremos como questão correta, aquela em que a conversão apresenta título coerente com o assunto abordado, coluna indicadora ou cabeçalho, número de linhas ou colunas corretas, fonte, além da estrutura dos dados em ordem decrescente. Quanto aos gráficos, esse deverá

conter unidades de análises utilizadas nas escalas, título ou legenda, bem como área proporcional às frequências de cada categoria, quando se tratar de gráfico de setor.

Partiremos para a análise do desempenho geral dos sujeitos na unidade de análise “Representação Semiótica em Estatística”.

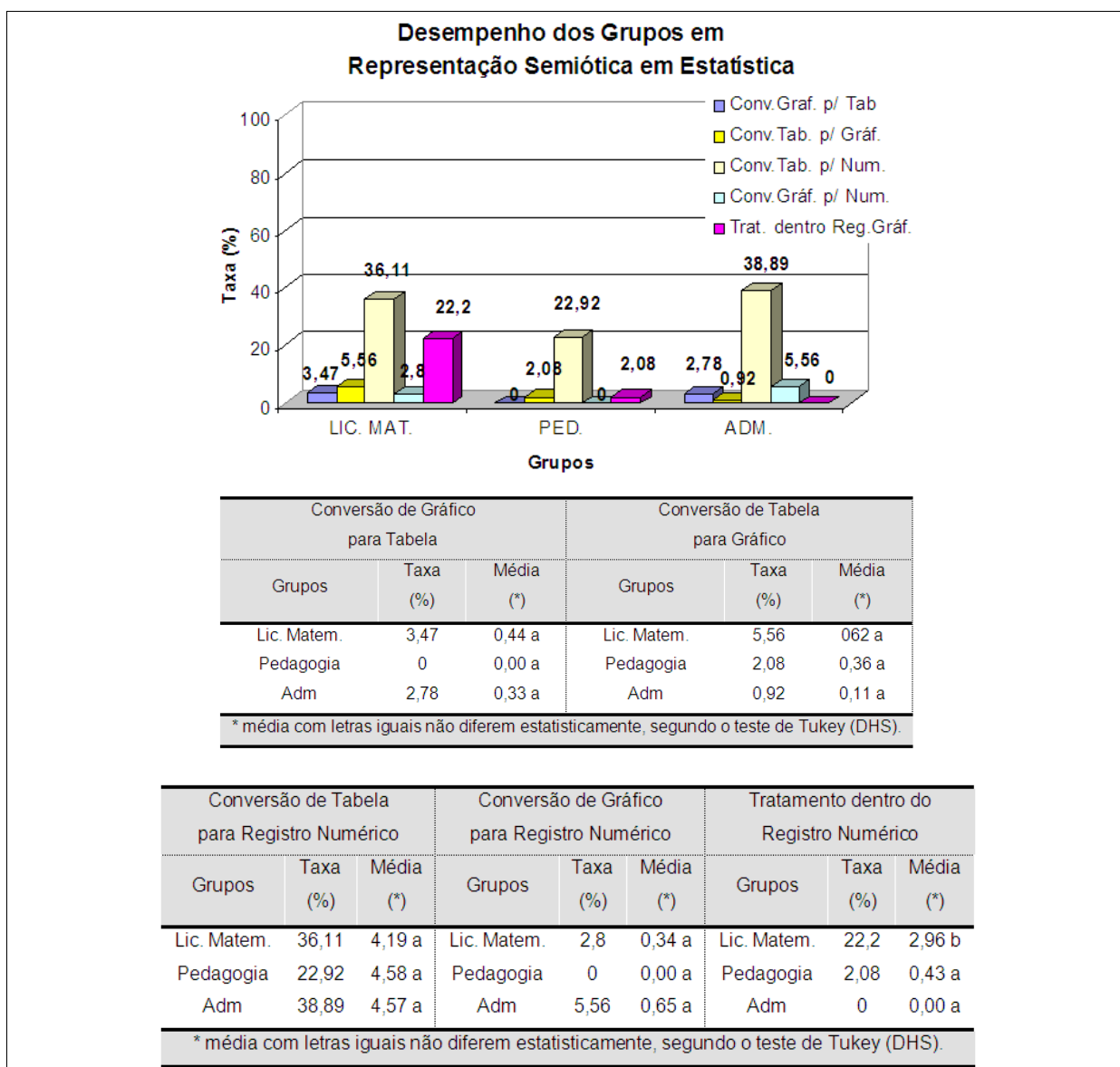


Figura 47: Desempenho dos grupos na unidade de análise Representação Semiótica em Estatística e tabelas com seus respectivos testes de Tukey.

O teste  $F$  indicou que os desempenhos dos sujeitos diferiram significativamente entre os grupos somente na categoria “tratamento dentro do registro gráfico” [ $F(2, 111) = 9,832; p = 0,000$ ].

Os itens que abordavam a “conversão de tabelas para registro numérico” representam as notas mais altas de todas as categorias e mesmo assim não passam de 39% de acertos. Nas categorias restantes, a taxa de acerto não ultrapassa 6%, com exceção dos sujeitos de Licenciatura em Matemática, que na categoria “tratamento dentro do registro gráfico”, atinge 22,2% de acertos.

Observamos neste caso que o desempenho nessa categoria foi, no mínimo, sofrível. Esse fato nos leva a crer que as conversões de um registro para outro não são tarefas tão simples, e que provavelmente o ensino desses conceitos não surta o efeito esperado.

Partiremos assim para a análise de cada uma das categorias.

### a) Conversão do Registro Gráfico para o Registro Tabular

Essa categoria contava com dois itens, Q.7 e Q.10.a. Discutiremos, individualmente, os itens que apresentam uma diferença significativa, segundo o teste qui-quadrado.

Tabela 15: Desempenho dos sujeitos, segundo os grupos, na “conversão do registro gráfico para o registro tabular”.

Questões	Grupos	Taxa de Acerto (%)			Teste qui-quadrado	
		Certo	Errado	Branco	$\chi^2(2)$	p-valor
Q.7	Lic. Mat.	1,4	70,8	27,8	92,346	<b>0,000</b>
	Pedag.	0	33,3	66,7		
	Adm.	3,7	63,0	33,3		
Q.10.a	Lic. Mat.	5,6	63,9	30,6	89,917	<b>0,000</b>
	Pedag.	0	33,3	66,7		
	Adm.	1,9	77,8	20,4		

Valores em negrito indicam que os desempenhos foram estatisticamente significativos segundo o teste qui-quadrado.

O item Q.7 trazia um gráfico de barras horizontais representando as respostas de crianças, por sexo, à retirada de um brinquedo. Solicitávamos que os sujeitos, com base na leitura dos dados contidos no gráfico, construíssem uma tabela. Nesse item, estamos considerando como conversão correta, aquelas tabelas que apresentam todos os elementos da construção: título, coluna indicadora, número de linhas e colunas, fonte e presença do valor total. Essa exigência se deu por estarmos trabalhando com alunos do Ensino Superior.

Os desempenhos dos sujeitos não diferiram estatisticamente entre os grupos na questão Q.7 [ $F(2, 103) = 0,771$ ;  $p = 0,465$ ]. O teste de Tukey (DHS) indicou que não houve diferença entre os desempenhos dos sujeitos ( $p = 0,444$ ).

Tabela 16: Desempenho dos sujeitos, segundo os grupos na questão Q.7 com respectivo teste de Tukey.

Categoria "conversão registro gráfico para registro tabular" – Q.7.		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	1,4	0,0192 a
Pedagogia	0	0,0000 a
Adm.	3,7	0,0556 a

\* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).

Percebemos que o desempenho dos sujeitos na conversão correta de um gráfico estratificado para uma tabela de dupla entrada com todos os seus elementos, não foi tão positivo. Apontamos essa baixa taxa de acertos, à questão da falta de algum elemento na construção das tabelas, como por exemplo, ao fato de apenas dois sujeitos dentre os 174 pesquisados, anotarem o título e fonte.

Observamos também, que a taxa de construções erradas para o GFLM e GFADM passa dos 50%, enquanto o GFP preferiu nem tentar fazer esta questão, ou não entendeu o enunciado do problema. Acreditamos que essa taxa, principalmente para os responsáveis pelo ensino da Estatística nas escolas, é no mínimo, preocupante.

Observe algumas construções consideradas erradas. Na figura 48, um sujeito do GFLM não anotou os elementos como título, as colunas indicadoras ou cabeçalho, fonte, valor total, além de manter uma coluna a mais.

MASCULINO	PRO. BRINQUEDO	15	5
	RETIRADA	(15/20)	5
	RAIJA		15
	CHORO		25
FEMININO	PRO. BRINQUEDO	5	15
	RETIRADA	(5/20)	3
	RAIJA	15	3
	CHORO		28

Figura 48: Exemplo de resposta de um sujeito do GFLM ao item Q.7.

Outro sujeito, do mesmo grupo (Figura 49), não anotou a coluna indicadora ou cabeçalho, o título, valor total e a fonte.

PRO. Brinquedo	5	15
Retirada	5	4
Raija	15	3
Choro	25	28
MASCULINO		FEMININO

Figura 49: Exemplo de resposta de um sujeito do GFLM ao item Q.7.

Já no GFADM encontramos sujeitos que não conseguiram entender o que era pedido na questão (Figura 50).

<p>nos entendi o que é pl fazer</p>		
-------------------------------------	--	--

Figura 50: Exemplo de resposta de um sujeito do GFADM ao item Q.7.

Outro sujeito, também do GADM, não anotou os elementos como título, fonte, valor total, além de deixar a coluna da variável nominal sem cabeçalho (Figura 51). Observe que nas células das colunas, o sujeito apontou os valores das variáveis nominais juntamente com o símbolo de porcentagem, e, no entanto, não calculou a taxa percentual.

pinguim	5%	Feminino	15%
vestido	5%		9%
vestido	15%		3%
chão	25%		28%

Figura 51: : Exemplo de resposta de um sujeito do GFADM ao item Q.7.

O item Q.10.a solicitava aos sujeitos que, construíssem uma tabela, a partir do registro figural de cinco bonecos com seus respectivos nomes e alturas. Essa construção também deveria apresentar todos os elementos que foram solicitados na Q.7, além de uma estrutura decrescente.

Tabela 17: Desempenho dos sujeitos, segundo os grupos no item Q.10.a com respectivo teste de Tukey.

Categoria "conversão registro gráfico para registro tabular" – Q.10.a		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	5,6	0,08 a
Pedagogia	0	0,00 a
Adm.	1,9	0,0233 a

\* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).

Os desempenhos dos sujeitos não diferiram estatisticamente entre os grupos na questão Q.10.a [ $F(2, 108) = 1,296$ ;  $p = 0,278$ ], além de serem, no mínimo sofríveis.

Quando fizemos a análise a priori desse item, esperávamos um bom desempenho dos sujeitos; no entanto, já apontávamos para a dificuldade em construir uma tabela com todos os seus elementos necessários.

Assim sendo, na Tabela 17, por meio de suas taxas de acertos, demonstra a dificuldade na construção apontada. Ressaltamos que foram consideradas como construção correta, aquelas que trazem todos os elementos já discutidos na questão anterior.

Encontramos os sujeitos justificando que já estudaram, mas que não conseguem lembrar como registrar a tabela ou que não conseguem concluir. Acreditamos que o tipo de ensino estatístico, nesse caso, não foi significativo, uma vez que os sujeitos não converteram do registro figural para o registro tabular.

Veja exemplos desses tipos de respostas, que apontam para esse fato, na Figura 52. Ainda, conjecturamos que provavelmente este ensino deficiente no quesito construção de tabelas se faz mais presente no GFP, haja vista a taxa de acertos desse grupo nesse item.

a) Com base na leitura dos dados acima, construa uma tabela. Para facilitar a construção tem-se abaixo uma grade. Utilize o quanto for necessário e risque o restante.

Bia	145			
Ana	148			
Caio	150			
Luz	152			
João	155			

Estudei -  
mas não  
lembro  
como  
fazer.

	Nome	Estatura	PA	PG	Total
1	Bia	145	145 x 100		
2	Ana	148			
3	Caio	150			
4	Luz	152	no conselho	ou seja	
5	João	155	no conselho	como faz	

ou cálculo

Figura 52: Exemplos de respostas de dois sujeitos do GFP ao item Q.10.a.

Passamos então, a analisar os tipos de erros cometidos por dois sujeitos do GFLM.

ALUNO	ALTURA(Com)
LUIZ	152
ANA	148
JOÃO	155
BIA	145
CAIO	150

Aluno	140-150	150-160
Bia	X	
Ana	X	
Luz		X
João		X
Caio		X

Figura 53: Exemplos de respostas de dois sujeitos do GFLM ao item Q.10.a.

Observamos que a estrutura da tabela (Figura 53) criada por estes dois sujeitos não apresentam título ou total, e, a estrutura não está em ordem decrescente; no exemplo da direita, além de tudo, não registrou o cabeçalho das células da tabela, deixou uma linha em branco e tentou construir uma tabela de distribuição dos cinco alunos, segundo a faixa etária.

### b) Conversão do Registro Tabular para o Registro Gráfico

Essa categoria também conta com dois itens, Q.4.3 e Q.8. Discutiremos os itens que apresentam uma diferença significativa, segundo o teste qui-quadrado, individualmente.

Tabela 18: Desempenho dos sujeitos, segundo os grupos, na “conversão do registro tabular para o registro gráfico”.

Conversão de Registro Tabular para Registro Gráfico	Questões	Grupos	Taxa de Acerto (%)			Teste qui-quadrado	
			Certo	Errado	Branco	$\chi^2(2)$	p-valor
Q.4.3.		Lic. Mat.	5,6	80,6	13,9	118,478	<b>0,000</b>
		Pedag.	0	58,3	41,7		
		Adm.	0	81,5	18,5		
Q.8.		Lic. Mat.	5,6	54,2	40,3	75,922	<b>0,000</b>
		Pedag.	4,2	41,7	54,2		
		Adm.	1,9	66,7	31,5		

Valores em negrito indicam que os desempenhos foram estatisticamente significativos segundo o teste qui-quadrado.

O item Q.4.3 solicitava aos sujeitos que convertessem os dados registrados em uma tabela simples de dados agrupados para um gráfico. Para a construção do gráfico, consideramos como corretas, aquelas que apresentassem as unidades utilizadas nas escalas, título ou legenda e que respeitassem a escala.

Tabela 19: Desempenho dos sujeitos, segundo os grupos no item Q.4.3 com respectivo teste de Tukey

Categoria "conversão registro gráfico para registro tabular" – Q.4.3.		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	4,0	0,0645 a
Pedagogia	0	0,00 a
Adm.	0	0,00 a

\* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).

Os desempenhos dos sujeitos não diferiram estatisticamente entre os grupos nessa questão [ $F(2, 108) = 1,296$ ;  $p = 0,278$ ]. Observando a taxa de itens corretos dos três grupos, podemos apontar que seus desempenhos foram, no mínimo preocupantes.

O grupo dos futuros Licenciados em Matemática obteve a maior taxa de acerto, 4%, o que significa um desempenho sofrível em uma turma que, eventualmente se sairia melhor. Ressaltamos, ainda, que estamos trabalhando com sujeitos no nível superior e, por esse motivo, avaliamos todos os elementos que são necessários para a construção de um gráfico, uma vez que a falta de algum já caracterizava erro.

Passaremos, agora, a analisar os tipos de erros cometidos. Na Figura 54 encontramos dois exemplos de “não conversão” do registro tabular para o registro gráfico: nos dois gráficos os sujeitos não anotaram as unidades utilizadas nas escalas, tanto na construção da esquerda como da direita. Além disso, na construção da direita, o sujeito também não respeita escala e não apresenta título ou legenda. Ressaltamos que foi oferecido aos sujeitos materiais para auxiliar a construção correta como régua.

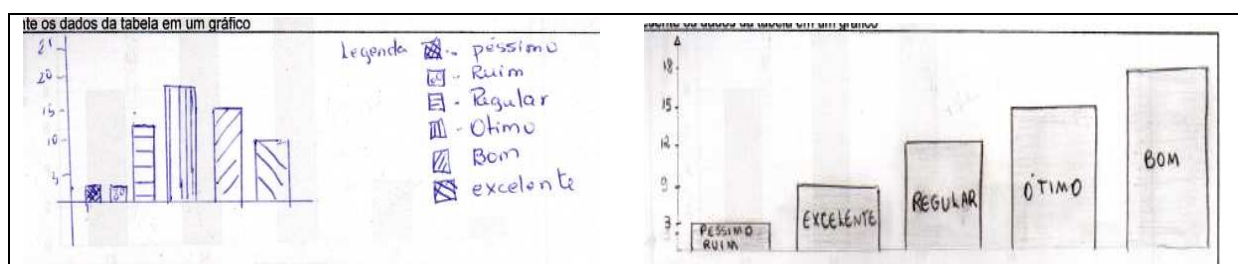


Figura 54: Exemplos de respostas de dois sujeitos do GFLM ao item Q.4.3.

Nos exemplos da Figura 55, encontramos na construção do gráfico da esquerda indícios de um gráfico cartesiano. No gráfico da direita, encontramos a utilização de um gráfico de linha que, costumeiramente é utilizado para demonstrar as modificações de uma variável em uma ou em um grupo de variáveis, ou ainda, a demonstração da evolução dessa variável durante algum período.

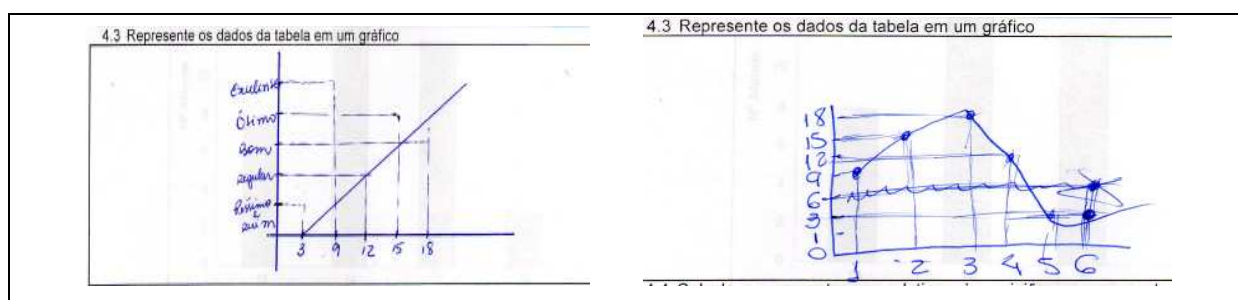


Figura 55: Exemplos de respostas dos sujeitos do GFADM à questão Q.4.3.

Na Figura 56, encontramos um exemplo de gráfico que não segue a estrutura tradicional da construção de um gráfico. O sujeito de GFLM construiu um triângulo com o vértice para baixo, dividiu em cinco partes (a tabela apresentava seis escalas de opinião, sendo que “ruim” e “péssimo” tiveram a mesma frequência) e classificou cada parte do triângulo comparando sua área com a frequência da opinião observada na tabela.

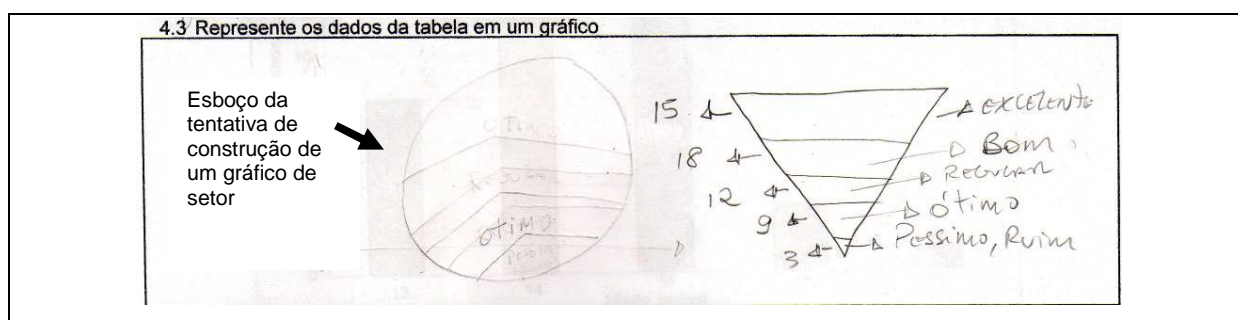


Figura 56: Exemplo de resposta de um sujeito do GFLM à questão Q.4.3.

Observe que na esquerda do exemplo da Figura 56, o mesmo sujeito esboçou uma tentativa de converter a tabela para gráfico de setor; no entanto, as divisões não partiam do centro da circunferência. Entendemos que é relevante discutir o fato dessa construção ter partido de um sujeito de Licenciatura em Matemática, porque

provavelmente, esse sujeito será responsável pela formação estatística de alunos, e não nos parece que hoje, esse indivíduo esteja preparado para tal façanha.

Já na Figura 57, encontramos um exemplo da construção de um gráfico, que provavelmente, sofreu influência do gráfico de barras empilhadas da questão Q.3.

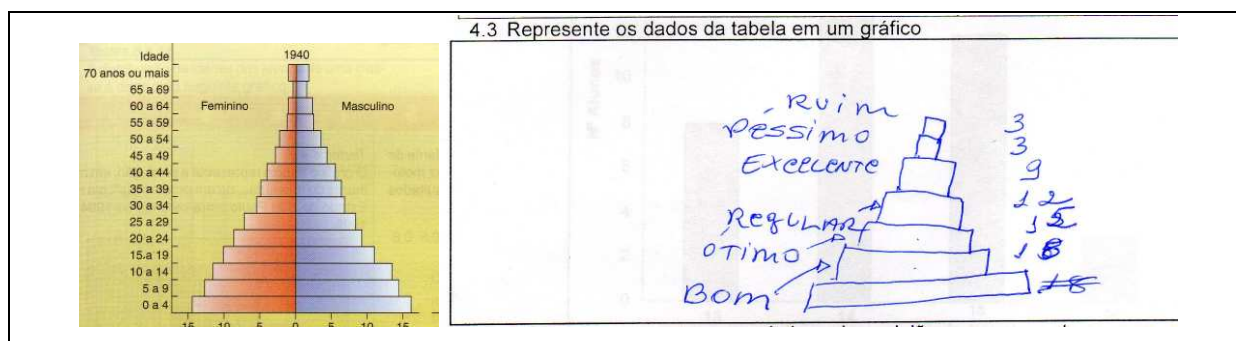


Figura 57: Exemplo de resposta de um sujeito do GFADM à questão Q.4.3.

No item Q.8, era solicitado que, segundo o gênero e a partir de uma tabela de distribuição de frequências da variável “uso do cinto de segurança”, os sujeitos construíssem um gráfico, por sexo, do entrevistado.

Os desempenhos dos sujeitos não diferiram estatisticamente entre os grupos tanto no teste de Tukey nessa questão, quanto no teste  $F [F(2, 101) = 0,775; p = 0,463]$ .

Tabela 20: Desempenho dos sujeitos, segundo os grupos, no item Q.8 e tabela com respectivo teste de Tukey.

Categoria “conversão registro tabular para registro gráfico” – Q.8.		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	5,6	0,093 a
Pedagogia	4,2	0,090 a
Adm.	1,9	0,027 a

\* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).

Na análise a priori, esperávamos um bom desempenho dos sujeitos nesse item, o que não ocorreu. Esse desempenho negativo se deve a não presença de os elementos da construção de um gráfico.

Vamos nos deter aos tipos de erros cometidos pelos sujeitos. A Figura 58 apresenta duas construções de gráficos considerados errados. Na construção da esquerda, o sujeito do GFLM não anotou as unidades utilizadas nas escalas, não respeitou escala e ainda construiu dois gráficos separados, um para o gênero feminino e outro para o masculino.

8. Analise as informações da tabela:

**TABELA 2.15** Tabulação cruzada do uso do cinto de segurança por sexo

Uso do cinto de segurança	Sexo do entrevistado		Total
	Masculino	Feminino	
Sempre	144	355	499
Na maioria das vezes	66	110	176
Algumas vezes	58	66	124
Raramente	39	44	83
Nunca	60	55	115
Total	367	630	997

Fonte: Adaptado de Levin, Jack. Fox, J.A; Estatística para ciência Humanas. São Paulo: Pearson Prentice Hall, 2004.

Agora com base na leitura da tabela acima construa um gráfico, por sexo, do entrevistado:

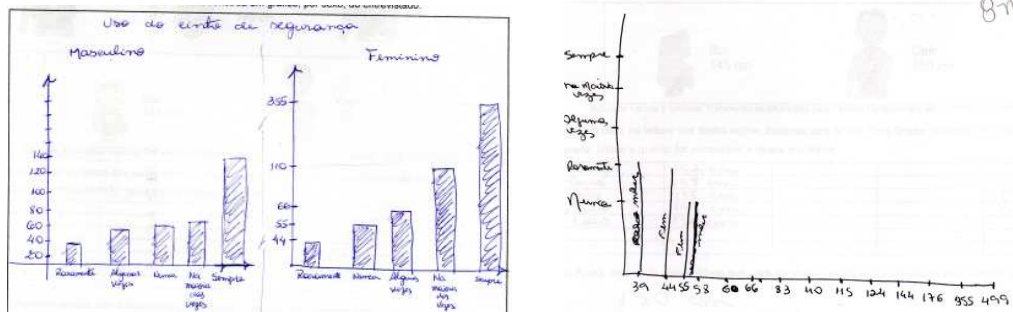


Figura 58: Exemplos de respostas dos sujeitos do GFLM e GFADM, respectivamente ao item Q.8.

Na construção do gráfico da direita, além de todos os elementos já ditos, o sujeito do GFADM iniciou a construção de um gráfico de barras, e, no entanto, partiu para a construção de um gráfico de colunas.

Na Figura 59, o sujeito do GFLM, no gráfico da esquerda, somou todas as frequências do gênero masculino e feminino, e logo após, construiu duas colunas com estas somas. Nota-se que ela não apresenta eixo, título, legenda ou escala.

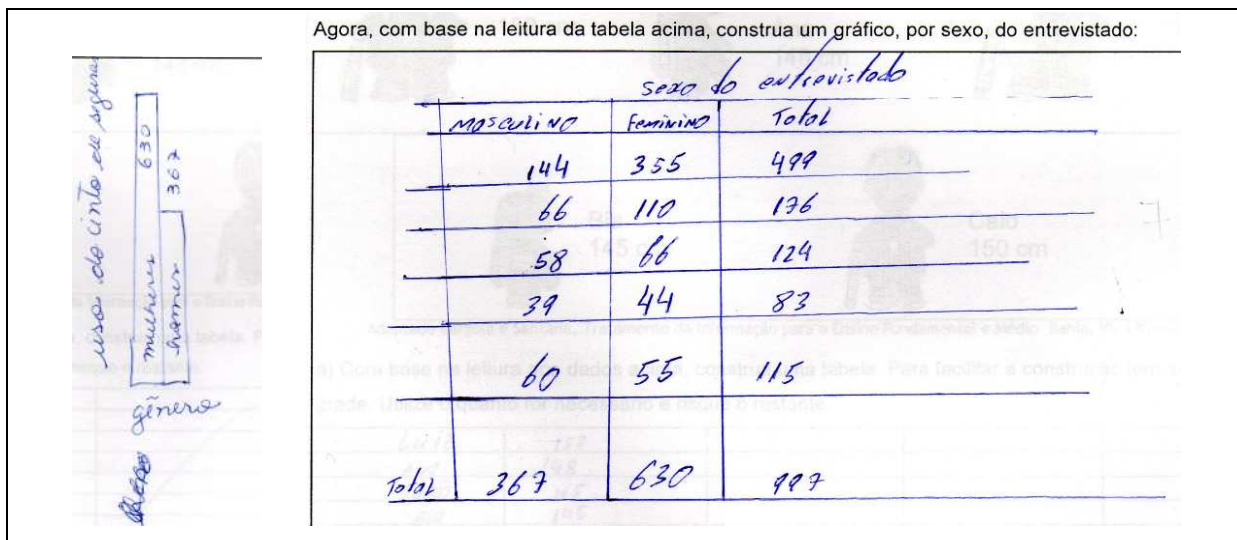


Figura 59: Exemplos de respostas dos sujeitos do GFADM e GFLM à questão Q.8, respectivamente.

Na mesma Figura 59, à direita, encontramos um sujeito do GFLM que, mesmo com o enunciado solicitando que construísse um gráfico, construiu uma tabulação.

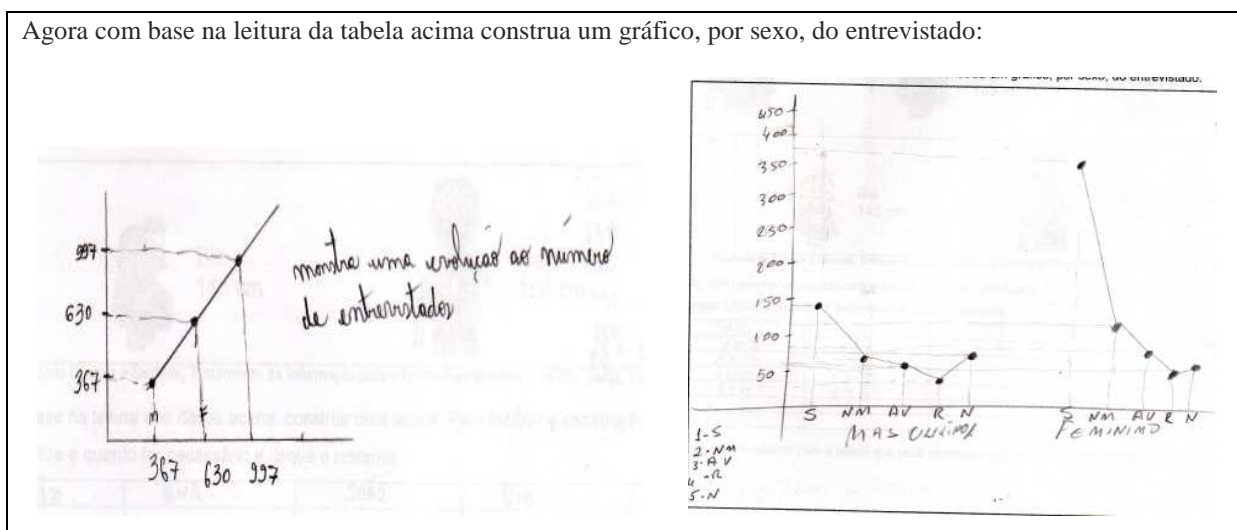


Figura 60: Exemplos de respostas dos sujeitos do GFADM e GFLM à questão Q.8, respectivamente

Na Figura 60, o gráfico da esquerda feito por um sujeito do GFADM, traz o esboço da construção de um gráfico cartesiano sem a presença das unidades utilizadas, do título ou legenda, e ainda, com os valores da escala retirados da linha da tabela onde constavam os valores totais. A construção do gráfico da direita, feito por um sujeito do GFLM, demonstra a construção de dois gráficos de linhas, que costumeiramente é utilizado para demonstrar as modificações de uma variável em

uma ou em um grupo de variáveis, ou ainda, demonstrar a evolução dessa variável durante algum período.

### c) Conversão do Registro Tabular para o Registro Numérico

Essa categoria contava com dois itens, Q.5.2 e Q.10.b. Apesar dos itens não apresentarem uma diferença significativa, segundo o teste qui-quadrado, promoveremos uma discussão individual de cada questão.

Tabela 21: Desempenho dos sujeitos, segundo os grupos, na “conversão do registro tabular para o registro numérico”.

Conversão de Registro Tabular para Registro Gráfico	Questões	Grupos	Taxa de Acerto (%)			Teste qui-quadrado	
			Certo	Errado	Branco	$\chi^2(2)$	p-valor
Q.5.2.		Lic. Mat.	41,7	37,5	20,8	0,508	<b>0,476</b>
		Pedag.	27,1	22,9	50,0		
		Adm.	44,4	38,9	16,7		
Q.10.b.		Lic. Mat.	30,6	16,7	52,8	3,568	<b>0,059</b>
		Pedag.	18,8	14,6	66,7		
		Adm.	33,3	24,1	42,6		

Valores em negrito indicam que os desempenhos foram estatisticamente significativos segundo o teste qui-quadrado.

Partiremos assim, para as análises individuais das questões.

O item Q.5.2 solicitava que a partir de uma tabela simples de dados agrupados, os alunos convertessem para o registro numérico “Moda”. É importante ressaltar que esse item já foi analisado sob a perspectiva da *Leitura e interpretação de Tabelas*, e, também nesse caso, os desempenhos não foram muito positivos.

Observamos que os desempenhos dos sujeitos não diferiram significativamente entre os grupos na questão Q.5.2 [ $F(2, 125) = 0,008$ ;  $p = 0,992$ ]. Tais resultados também foram diagnosticados no teste Tukey.

Tabela 22: Tabela com respectivo teste de Tukey e desempenho dos sujeitos, segundo os grupos, no item Q.5.2.

Categoria "conversão registro tabular para registro numérico" – Q.5.2.		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	41,7	0,53 a
Pedagogia	27,1	0,54 a
Adm.	44,4	0,53 a

\* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).

O problema do desempenho não tão positivo nesse item foi ocasionado porque os sujeitos trocaram a resposta. Queríamos que os sujeitos respondessem qual era a Moda da questão mais respondida, logo, eles deveriam dar como resposta “questão 2”, a variável independente; no entanto eles davam como resposta o número “40”, a variável dependente. Esse fato sinaliza que os sujeitos não convertem os dados agrupados de uma tabela simples para o registro numérico nas medidas de tendência central.

O item Q.10 dividia-se em duas partes: na primeira, solicitávamos que os sujeitos, a partir de um registro pictórico de cinco bonecos, que representavam alunos de uma série fictícia, seus nomes e respectivas alturas, construíssem uma tabela. Na segunda, que os sujeitos, a partir da tabela que eles construíram, calculassem a altura mediana dos meninos. Consideramos como valor correto, tanto o cálculo efetuado somente com o gênero masculino (menino) quanto com o total (todos os meninos), devido a dupla interpretação da palavra “meninos”.

Tabela 23: Tabela com respectivo teste de Tukey e desempenho dos sujeitos, segundo os grupos, no item Q.5.2.

Categoria "conversão registro tabular para registro numérico" – Q.10 b.		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	30,6	0,65 a
Pedagogia	18,8	0,56 a
Adm.	33,3	0,58 a

\* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).

Os desempenhos dos sujeitos não diferiram estatisticamente entre os grupos no item Q.10.b [ $F(2, 80) = 0,218$ ;  $p = 0,805$ ].

Observamos também que a taxa de questões em branco foi alta em todos os três grupos em detrimento da taxa de questões corretas (GFLM 52,78%, GFP 66,67% e GFADM 42,59%). Para o GFP e GFADM, chamou-nos a atenção o fato de alguns dos sujeitos se justificarem por ter deixado o item Q.10.b em branco, uma vez que não encontramos no GFLM nenhuma justificativa para tantas questões em branco.

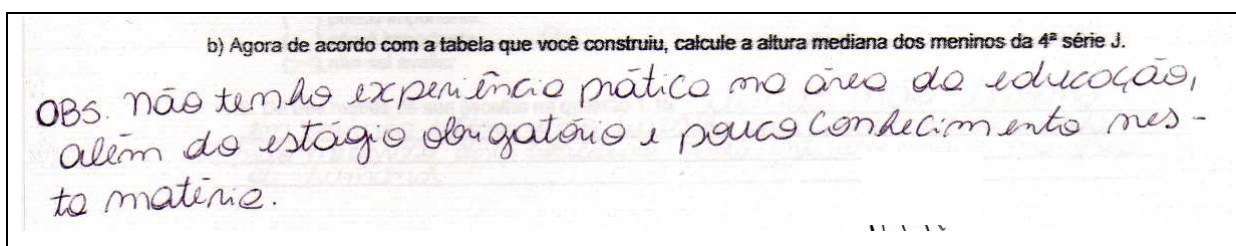


Figura 61: Exemplo de resposta de um sujeito do GFP para a questão Q.10.b.

Esse item não exigia um custo cognitivo alto; por este motivo, a expectativa quando da análise a priori, era que os sujeitos fossem bem sucedidos, o que não ocorreu.

Quanto às taxas de questões erradas, levantamos alguns tipos de erros na tentativa de explicar o porquê do desempenho insatisfatório. São eles:

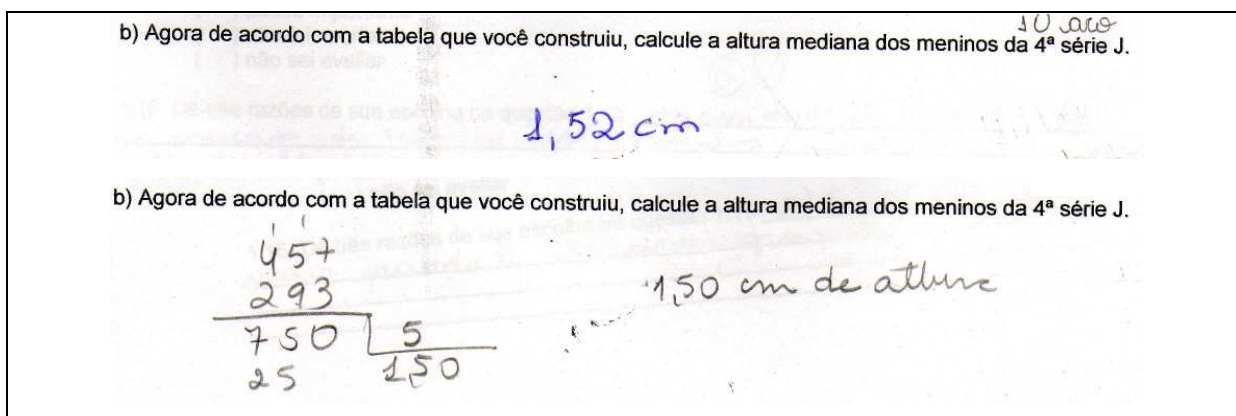


Figura 62: Exemplos de respostas de dois sujeitos do GFLM ao item Q.10.b

Observe que no primeiro exemplo da Figura 62, o sujeito não levou em consideração a unidade de medida. Provavelmente, ele ordenou os bonecos do

gênero masculino e observou qual a altura do boneco “do meio” (Luiz com 152 cm). Este sujeito quis transformar a unidade de medida de centímetros para metros, no entanto continuou anotando sua resposta em centímetro.

Já no segundo exemplo, o sujeito trata a mediana erroneamente, como se fosse média. Somou todas as alturas e dividiu pelo número de alunos, além de anotar a unidade de medida como o sujeito do exemplo anterior.

#### d) Conversão do Registro Gráfico para o Registro Numérico

Essa categoria contava com apenas um item, o Q.6.3. Ressaltamos que esse item já foi analisado sob a perspectiva de *Leitura e interpretação de gráficos*, no entanto, também se enquadra na conversão de um registro gráfico para o registro numérico (medidas de tendência central – mediana).

Os desempenhos dos sujeitos não diferiram estatisticamente entre os grupos no item Q.6.3, tanto no teste  $F$  [ $F(2, 128) = 0,942; p = 0,393$ ], como no teste de Tukey. No entanto, houve diferença significativa nos desempenhos dos sujeitos nessa questão, segundo o teste qui-quadrado [ $\chi^2(1, N=129) = 109,775; p = 0,000$ ].

O item em discussão solicitava que os sujeitos calculassem a idade que representa a mediana de um gráfico de barras representando a distribuição de idades de alunos fictícios.

Tabela 24: Tabela com respectivo teste de Tukey e desempenho dos sujeitos, segundo os grupos, no item Q.6.3.

Categoria “conversão registro gráfico para registro numérico” – Q.6.3.		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	2,8	0,0345 a
Pedagogia	0	0,0000 a
Adm.	5,6	0,0652 a

\* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).

Como podemos observar na Tabela 23, os desempenhos dos grupos não foi positivo nessa questão. O problema detectado para um desempenho tão baixo foi que provavelmente, os sujeitos não atentaram para o fato de que o gráfico representava uma distribuição de idades. Então, calcularam a Mediana pela “altura da coluna do gráfico” ou pela frequência, o que nos dois casos tornaram as respostas incorretas.

### e) Tratamento dentro do Registro Gráfico

Essa categoria continha somente um item, o Q.4.4. Ele solicitava que, a partir do gráfico da questão Q.4.3 construída pelo próprio sujeito, ele calculasse as porcentagens relativas às opiniões e representasse-as num gráfico de setores. Para tal construção, os sujeitos já encontravam uma circunferência dividida de 5 em 5°. Essa medida visava facilitar tal construção.

Tabela 25: Tabela com respectivo teste de Tukey e desempenho dos sujeitos, segundo os grupos, no item Q.4.4.

Categoria “tratamento dentro do registro gráfico” – Q.4.4.		
Grupos	Taxa (%)	Média (*)
Lic. Matem.	22,2	0,30 b
Pedagogia	2,1	0,0435 a
Adm.	0	0,00 a
* média com letras iguais não diferem estatisticamente, segundo o teste de Tukey (DHS).		

Na análise a priori já esperávamos um baixo índice de acerto, uma vez que são necessários muitos “passos” de conversão para chegar à resolução. Estamos falando de uma conversão tripla (registro gráfico para registro numérico de porcentagem, depois para registro numérico em graus) e um tratamento dentro do registro gráfico.

Queremos ressaltar, que para esse item ser considerado correto, analisamos se a construção do gráfico apresentava área proporcional às frequências de cada categoria.

Conforme prevíamos numa análise anterior, o desempenho dos sujeitos não foi tão positivo, salvo o GFLM que conseguiram a maior taxa de acerto quando comparado aos outros dois grupos. Apoiados nos resultados do teste de Tukey e  $F$  [ $F(2, 111) = 9,832; p = 0,000$ ], constatamos que foram os desempenhos dos grupos GFP e GFADM responsáveis pela diferença significativa entre os resultados dos grupos.

Acreditamos que essa diferença foi causada pelo fato dos sujeitos de Licenciatura em Matemática, apresentar mais habilidade em transformar valores em porcentagem, e, porcentagem em graus. No entanto, o GFLM apresenta mais de 50% dos pesquisados errando essa questão por algum motivo.

Nesse sentido, partiremos para analisar os tipos de erros cometidos pelos sujeitos.

Nota-se no exemplo de resposta do sujeito do GFLM (Figura 63), que este não apresenta a habilidade em transformar porcentagem em graus.

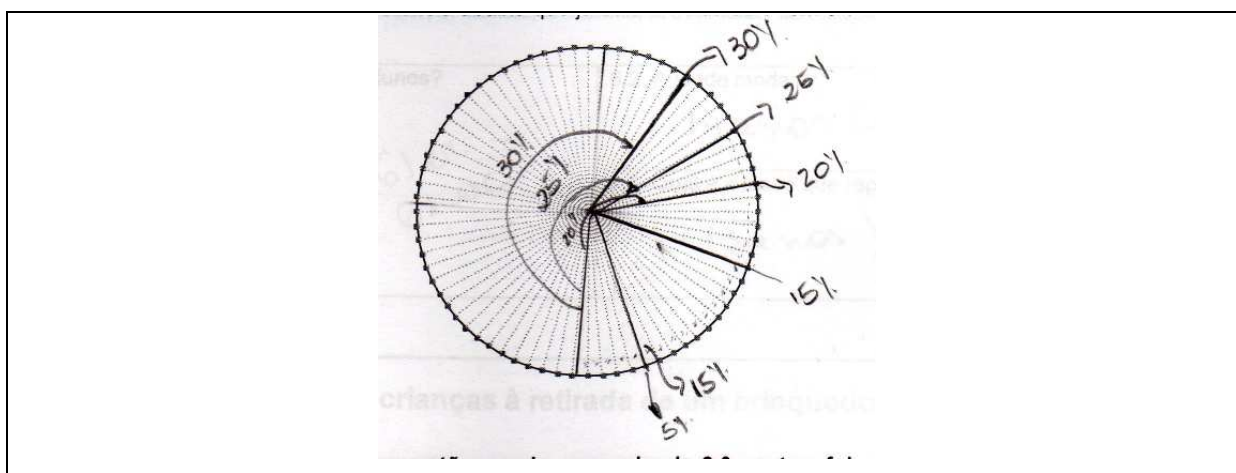


Figura 63: Exemplo de resposta de um sujeito do GFLM ao item Q.4.4.

Já na Figura 64, o sujeito do GFADM apresenta dificuldade em transformar a porcentagem em graus, o que nos leva a concluir que tal sujeito tentou trabalhar com a divisão do gráfico em graus e, provavelmente por dificuldade, abandonou a construção e partiu para uma nova construção com os valores em porcentagem.

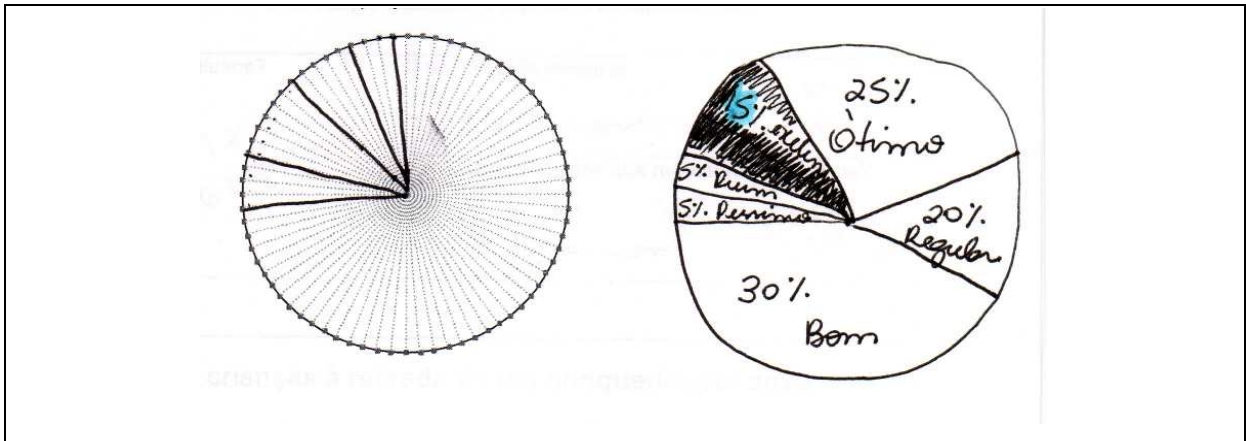


Figura 64: Exemplo de resposta de um sujeito do GFADM ao item Q.4.4.

Quanto às taxas de itens em branco, principalmente nos sujeitos do GFP, supomos que este item exigia um alto custo cognitivo. Observe a justificativa de um dos sujeitos do GFP por não ter resolvido o item Q.4.4.

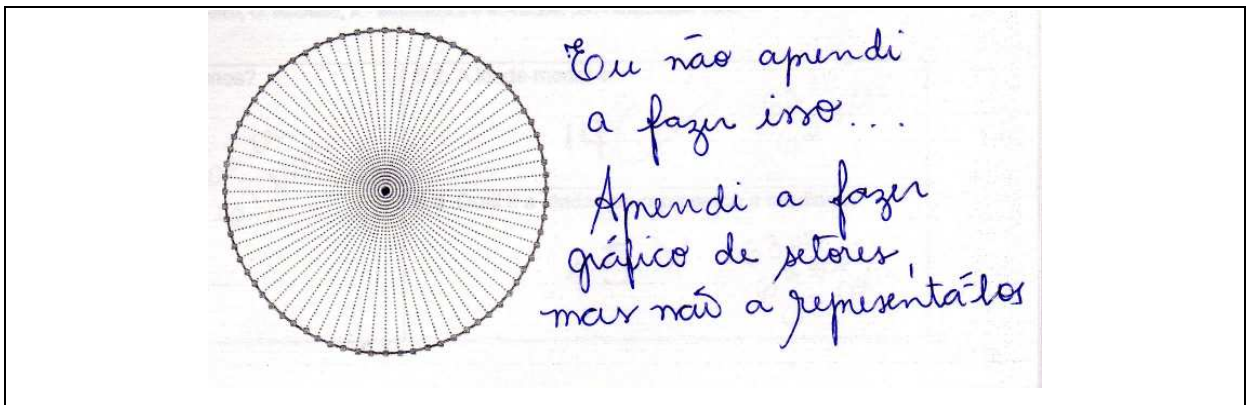


Figura 65: Exemplo de resposta de um sujeito do GFP ao item Q.4.4.

#### 4.1.2.4. O DESEMPENHO EM CÁLCULO DE MODA, MÉDIA E MEDIANA

Esta análise tem caráter secundário, uma vez que não é o foco principal de nossa pesquisa. No entanto, é importante ressaltar os desempenhos dos sujeitos pesquisados nos cálculos de Moda, Média e Mediana, uma vez que, estas são as medidas de tendência centrais mais elementares da Estatística.

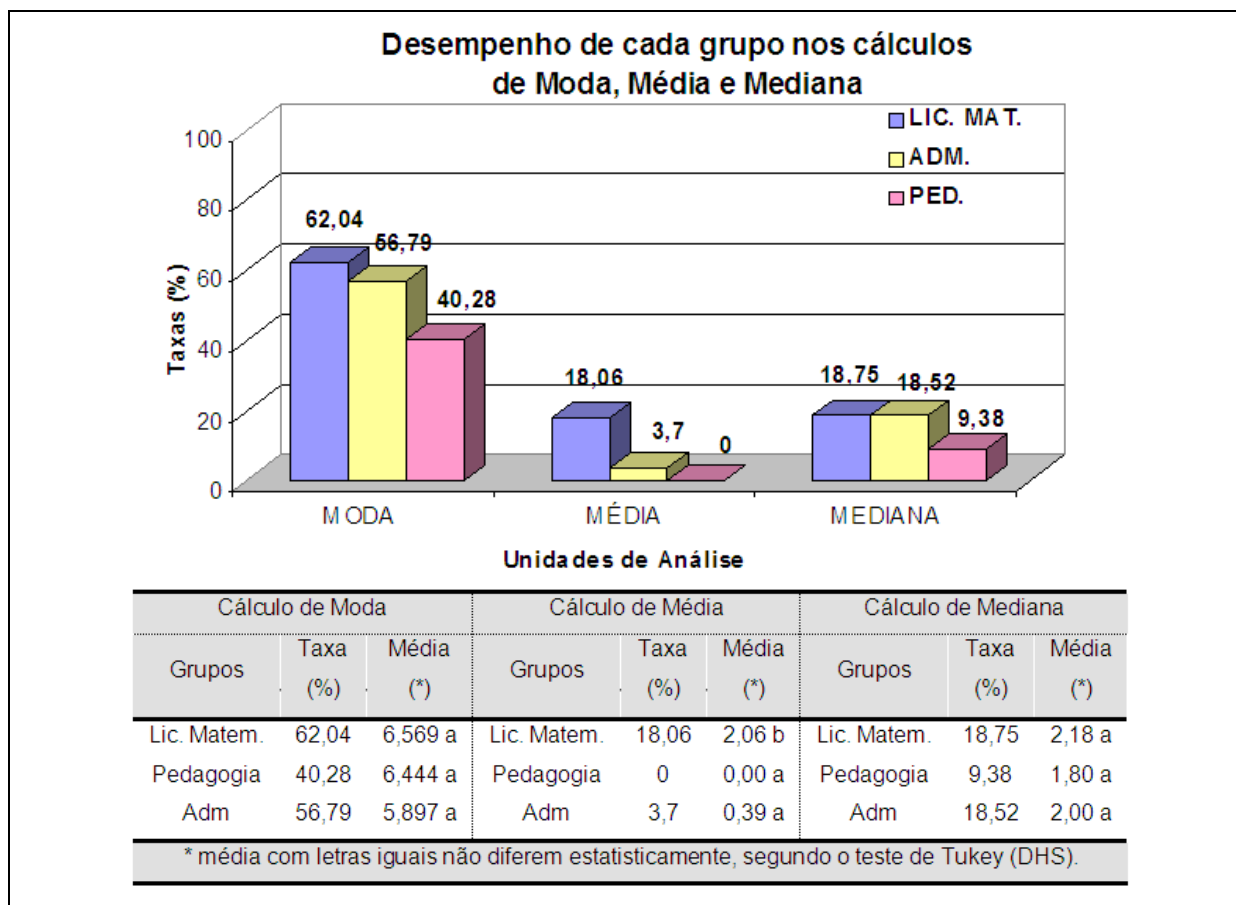


Figura 66: Desempenho dos grupos em cálculos de moda, média, mediana e tabelas com seus respectivos testes de Tukey.

Os desempenhos dos sujeitos não diferiram estatisticamente entre os grupos nas questões que abordavam Moda e Mediana [ $F(2, 149) = 0,846$ ;  $p = 0,431$ ], [ $F(2, 136) = 0,195$ ;  $p = 0,823$ ], respectivamente. Já nas questões que abordavam o cálculo da Média, os desempenhos dos sujeitos diferiram significativamente [ $F(2, 138) = 6,346$ ;  $p = 0,002$ ].

Como podemos observar no gráfico da Figura 66, os sujeitos se saíram melhor nas questões que abordam a Moda, seguido do cálculo da Mediana e do cálculo da Média.

Apontamos como provável facilitador do bom desempenho nas questões que abordavam a Moda os níveis das questões propostas. Tais questões não exigiam alto custo cognitivo; demandavam, no máximo, a interpretação dos dados e o uso de outras habilidades matemáticas como adicionar, subtrair, multiplicar e dividir os dados para comparar quantidades, combinar e integrar as informações contidas no gráfico.

Também apontamos a leitura e a interpretação adequadas do gráfico de distribuição como provável causa do baixo desempenho no cálculo da Média.

Quanto ao cálculo da Mediana, o problema maior foi ler e interpretar adequadamente o gráfico da distribuição que estava incluído no enunciado do problema, além do cálculo da Mediana a partir de um registro figural.

### TECENDO AS CONSIDERAÇÕES FINAIS

---

A realização do presente estudo teve por objetivo investigar os conhecimentos básicos dos alunos dos cursos de Licenciatura em Matemática, Pedagogia e bacharelado em Administração com relação à leitura e a interpretação de gráficos e tabelas, estudados na disciplina de Estatística. Especificamente, realizamos uma pesquisa para responder à seguinte questão:

QUAIS SÃO OS CONHECIMENTOS SOBRE LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS QUE ALUNOS DOS CURSOS DE LICENCIATURA EM MATEMÁTICA, PEDAGOGIA E BACHARELADO EM ADMINISTRAÇÃO MOBILIZAM AO RESOLVER SITUAÇÕES APRESENTADAS EM FORMA DE PROBLEMAS?

Para alcançarmos o objetivo proposto e termos subsídios suficientes para responder a nossa questão de pesquisa traçamos um estudo, constituído de cinco capítulos. Neste capítulo, intitulado “Tecendo as Considerações Finais”, será feito o resgate do percurso traçado ao longo do estudo. Na sequência, será apresentada a síntese dos resultados para, a partir dela, respondermos a nossa questão de pesquisa. Por fim, encaminharemos algumas sugestões que possam contribuir a futuras pesquisas.

#### 5.1. RESGATE DO PERCURSO DA PESQUISA

Quando do início desta dissertação (Capítulo I) promovemos a discussão de como as informações Estatísticas chegam às pessoas e, ainda, de como estas são

trabalhadas nas escolas. Nesse sentido, concluímos que seria desejável que os educadores dominassem minimamente a Estatística básica por ela fazer parte do bloco de conteúdo ou tema estruturador a ser ensinado em tais escolas.

Tal argumento justificou, portanto, a opção em pesquisar como a Estatística é ensinada no Ensino Superior, em específico a alunos dos cursos de Licenciaturas em Matemática e em Pedagogia. Para a escolha dos cursos, consideramos o fato de os alunos serem os futuros responsáveis por difundir a Estatística que aprenderam aos estudantes do ensino básico. Concomitantemente, investigamos o ensino da Estatística em um curso que a utiliza num contexto de usuário, como o de Administração. Foi realizado o estudo comparativo entre os conhecimentos básicos dos alunos dos cursos, em leitura e interpretação de gráficos e tabelas.

Na sequência, apresentamos uma sucinta introdução aos conceitos básicos estatísticos como cálculo de Moda, Mediana e Média e discutimos acerca da Educação Estatística e de algumas idéias que permeiam esta questão, como o Letramento e o Raciocínio Estatístico (Capítulo II).

Em seguida, para que pudéssemos estruturar nossa pesquisa, foi necessário contar com alguns aportes teóricos, também no Capítulo II. Dessa forma, e com relação à leitura e compreensão dos dados gráficos, apoiamos-nos na teoria de Frances R. Curcio (1989), que define três níveis de compreensão gráfica: leitura dos dados, leitura entre os dados e leitura além dos dados. Para a leitura e a interpretação de tabelas nos apoiamos na estrutura teórica das representações tabulares de Howard Wainer (1992), classificadas em níveis básico, intermediário e avançado. Houve fundamentação, ainda, na teoria de Registro de Representação Semiótica de Duval (2005), por meio da qual analisamos os resultados da perspectiva da conversão e tratamento.

Amparados nos aportes teóricos, desenvolvemos a pesquisa descritiva com coleta do tipo naturalista ou de campo, cujo objeto de estudo é colhido diretamente em ambiente, sem intervenção ou manuseio do pesquisador (SEVERINO, 2008, p. 123).

Participaram da pesquisa 174 sujeitos, divididos da seguinte forma: 72 sujeitos de Licenciatura em Matemática, 48 sujeitos de Pedagogia e 54 de bacharelado em Administração (Capítulo III). Responderam a um teste diagnóstico

composto de dez questões, perfazendo 24 itens analisados. O resultado do teste aplicado foi discutido no Capítulo IV.

Por ser de suma relevância para podermos responder à questão de pesquisa de nosso estudo, faremos uma síntese desses resultados na seção a seguir.

O presente Capítulo pretende fazer o fechamento de nossa pesquisa, apresentando os resultados apurados no teste aplicado e já discutido no Capítulo IV.

Para que as considerações sejam melhor apresentadas dividiremos este capítulo em três partes: na primeira será abordada uma síntese dos principais resultados obtidos; na segunda, responderemos a nossa questão de pesquisa e, na terceira, apontaremos algumas sugestões para futuras pesquisas.

## **5.2. SÍNTESE DOS PRINCIPAIS RESULTADOS OBTIDOS**

Nessa seção, apresentaremos a síntese dos resultados em duas etapas. Na primeira, será demonstrado o desempenho geral dos sujeitos no teste, segundo os grupos. Na segunda, o desempenho dos sujeitos em cada unidade de análise de acordo com os grupos, além da análise do desempenho – em caráter secundário - dos sujeitos em cálculo de Moda, Média e Mediana.

### **5.2.1. SOBRE O DESEMPENHO GERAL NO TESTE**

Nas questões, quando analisamos o desempenho dos grupos como um todo, constatamos que os sujeitos do grupo futuros licenciados em Matemática tiveram desempenho mais satisfatório do que o dos futuros Administradores. Os dois últimos, estes dois grupos por sua vez, saíram-se de forma mais positiva em relação aos futuros Pedagogos. Apesar da melhor performance, entretanto, os melhores escores não superam 50% de acertos.

É importante ressaltar que a taxa de questões em branco, no grupo dos futuros Pedagogos, foi maior que a taxa de acertos. Apontamos como provável causa para essa atitude de não responder as questões, a falta de motivação dos sujeitos ou uma insegurança na resolução do teste que abordava conteúdos estatísticos considerados básicos, além da leitura e da interpretação de gráficos e tabelas.

Estes resultados e atitudes reforçam as ideias de Grácio, Oliveira e Oliveira (s.d.). Os autores apontam que o trabalho com Estatística como “disciplina de serviço” se torna significativo se a ótica de investigação for sugerida pelos alunos. Com isso, sentimo-nos inclinados a inferir que, para este grupo de sujeitos, o ensino da Estatística pode ter sido ministrado sem os preceitos da análise exploratória dos dados e, por esse motivo, justifica-se o desempenho abaixo das expectativas.

### **5.2.2. SOBRE O DESEMPENHO NAS UNIDADES DE ANÁLISE**

Nessa seção iremos traçar o comparativo entre os desempenhos dos grupos nas questões sobre Leitura e Interpretação de Gráficos, Leitura e Interpretação de Tabelas e Representação Semiótica em Estatística.

Muito embora nossos dados tenham sido retirados de uma população de faculdades ou universidades particulares que representa a maioria dos estudantes brasileiros de nível superior, não possuímos dados estatísticos suficientes que nos permitam inferir para além de nossa população.

Na unidade *Leitura e Interpretação de Gráficos*, os grupos dos futuros licenciados em Matemática e Administradores foram significativamente mais positivos do que o desempenho dos futuros Pedagogos. Em outras palavras, quando analisamos as taxas de acertos dos futuros licenciados em Matemática e Administradores observamos que foram bem próximas, o que demonstra certa homogeneidade entre o desempenho dos grupos.

Por outro lado, essa homogeneidade ganha outros contornos quando analisado o desempenho dos dois grupos nas categorias da unidade. Na categoria

“nível leitura dos dados” essa homogeneidade é mantida; no entanto, na categoria “nível leitura entre os dados” e “nível leitura além dos dados” os desempenhos diferiram entre os grupos.

Para o nível “leitura entre os dados” os desempenhos dos três grupos não diferiram estatisticamente, sendo que o item Q.6.3 (leitura de um gráfico de coluna que representava a distribuição de idades de alunos fictícios) não foi um agente facilitador para os três grupos. As taxas de acerto para essa categoria não passam dos 7%, enquanto a taxa de questões em branco para o grupo dos futuros Pedagogos ultrapassa 40%.

Para o nível “leitura além dos dados” temos os piores desempenhos desta unidade de análise. O grupo dos futuros licenciados em Matemática obteve o melhor índice, cerca de 20%, enquanto o grupo dos futuros Pedagogos nem chegaram a pontuar. Novamente, temos a leitura de gráfico de colunas da distribuição de idade de alunos fictícios como um agente dificultador do desempenho dos três grupos.

Os desempenhos dos grupos nas categorias de *Leitura e Interpretação de Gráficos* seguem os preceitos da teoria de Curcio (1989) de que quanto maior a dificuldade atingida nos níveis de compreensão gráfica, menor é o desempenho.

Na unidade *Leitura e Interpretação de Tabelas* o grupo dos futuros licenciados em Matemática foi melhor sucedido do que o grupo de futuros Administradores e Pedagogos juntos, apesar de não passar de 40% de acerto. Tal desempenho vem ao encontro das ideias de Duval (2003), quando aponta para a importância de os adultos dominarem a leitura e a interpretação de tabelas. Existe a crença na simplicidade de comunicação de uma tabela, no entanto, os resultados de nossa pesquisa mostram que tal crença pode não ser verdadeira.

Nessa unidade temos os três grupos com desempenhos homogêneos na categoria “nível básico”, segundo os testes estatísticos. Apontamos como agente facilitador dessa categoria a tabela simples. É importante ressaltar que esse tipo de tabela significa uma das formas mais simples de representar as informações por serem utilizadas não só em livros ou artigos, mas também na publicidade. No entanto, apesar da pouca exigência cognitiva para a leitura, as taxas de acerto não ultrapassam de 50%.

A performance dos sujeitos nas questões no “nível intermediário” diferiram estatisticamente: o desempenho dos sujeitos do grupo dos futuros licenciados em Matemática foi mais positivo que os desempenhos dos outros dois grupos, apesar de não ultrapassar a taxa de 50% de acerto. Essa categoria exigia, segundo Wainer (1982), percepção da relação entre os dados e apontamos que essa percepção não foi um agente facilitador no desempenho dos sujeitos.

No “nível avançado”, os desempenhos dos sujeitos dos três grupos foram insatisfatórios, com taxa de acerto que não ultrapassa 30%. Estatisticamente, temos desempenho mais positivo dos sujeitos de licenciatura em Matemática do que o dos grupos de futuros Administradores e Pedagogos juntos. Além disso, observamos que as questões que exigiam maior entendimento das estruturas dos dados em sua totalidade se tornam um agente desfavorável ao desempenho dos sujeitos.

Para essa unidade de análise, também encontramos os resultados atingidos pelos sujeitos dos grupos seguindo os preceitos da teoria de Wainer (1982) de que quanto maior a exigência atingida nos níveis tabulares, menor é o desempenho.

Na unidade *Representação Semiótica em Estatística*, os desempenhos dos sujeitos dos três grupos não diferiram estatisticamente em quase todas as categorias, salvo o desempenho dos futuros licenciados em Matemática que se destacou na categoria “tratamento dentro do registro gráfico” e dos futuros Administradores na categoria “conversão de tabelas para registro numérico”.

Apuramos que tais desempenhos são, no mínimo, insatisfatórios. A maior taxa de acerto, pertence ao grupo dos futuros Administradores, na categoria “conversão de tabelas para registro numérico” e não passa dos 40%.

Como segunda maior taxa de acertos, temos o desempenho dos sujeitos de licenciatura em Matemática na categoria “tratamento dentro do registro gráfico”, com cerca de 25% de acerto, inclusive, a única categoria que indicou uma diferença nos desempenhos dos sujeitos, segundo os teste de Tukey. Consideramos essa taxa muito baixa, principalmente para aqueles sujeitos que, eventualmente, apresentam um arcabouço matemático subjacente ao estatístico.

Como discutido no Capítulo II, as representações semiótica na manipulação da Matemática é comumente considerada como a exteriorização das representações mentais, porém Duval (1993, apud DAMM, 1999) classifica esta visão como ingênu

e superficial, pois as representações semióticas são também necessárias para as atividades cognitivas do pensamento, em outras palavras, sem as representações semióticas o sujeito não constrói o conhecimento matemático aprendido. Nesse sentido, e analisando os resultados obtidos, podemos inferir que para esse grupo, a apreensão ou a produção de uma representação semiótica - “*semiósis*” – não foi internalizada e conseqüentemente a apreensão conceitual do objeto – “*noésis*” – também não, já que um não pode ser separado do outro.

Por não ser foco de nossa pesquisa, explanaremos os desempenhos dos sujeitos em cálculos de Moda, Média e Mediana, em caráter secundário. Os desempenhos dos sujeitos dos três grupos, não diferiram estatisticamente nos cálculos de Moda e Mediana e por esse motivo são homogêneos. Por outro lado, nas questões que abordavam o conceito de Média, os sujeitos do grupo de licenciatura em Matemática tiveram um desempenho mais positivo que os outros dois grupos.

No entanto, achamos relevante o fato da taxa de acertos do grupo dos futuros Administradores serem muito próximas da taxa de acerto dos futuros licenciados em Matemática em cálculo da Moda (cerca de 70%) e especialmente em Mediana (cerca de 20%). Tais taxas e equivalências nas questões que envolviam o conceito de Mediana, não eram esperadas. Acreditávamos que fossem estas questões bem mais trabalhadas nos cursos e por esse motivo, que obtivessem um melhor desempenho, principalmente no grupo dos futuros licenciados em Matemática.

### **5.3. RESPONDENDO A QUESTÃO DE PESQUISA**

Diante dos resultados apurados de nosso teste diagnóstico e de toda a discussão a cerca da leitura e interpretação de gráficos e tabelas e das representações semióticas em Estatística, partiremos agora para responder nossa questão de pesquisa:

*QUAIS SÃO OS CONHECIMENTOS SOBRE LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS QUE ALUNOS DOS CURSOS DE LICENCIATURA EM MATEMÁTICA, PEDAGOGIA*

*E BACHARELADO EM ADMINISTRAÇÃO MOBILIZAM AO RESOLVER SITUAÇÕES APRESENTADAS EM FORMA DE PROBLEMAS?*

Para responder a esta questão principal, elaboramos cinco outras questões, de caráter mais específico, cujas respostas contribuirão para que tenhamos uma compreensão mais sólida e abrangente dos conhecimentos básicos estatísticos dos grupos pesquisados.

No entanto, queremos explicar que para aquelas taxas menores que 20% de acertos, denominaremos o desempenho como “sofrível”; para acertos entre 25% e 50% “insatisfatório” e maiores que 50% como “satisfatório”. São elas:

- **Qual é o nível de compreensão gráfica que o grupo dos futuros licenciados em Matemática (GFLM), o grupo dos futuros Pedagogos (GFP) e o grupo dos futuros Administradores (GFADM) apresentam?**

Baseados em Curcio (1989) que, em uma revisão do trabalho de Bertin (1967, apud CAZORLA, 2002, p.54), definiu três níveis de compreensão gráfica, pudemos apurar por meio das taxas de acerto no teste que os três grupos atingiram satisfatoriamente apenas o nível de compreensão gráfica “Leitura dos dados”.

Para Curcio (1989) esse é o nível de compreensão em que o indivíduo é capaz de compreender somente os fatos explícitos que observa nos dados, não existindo interpretação e o nível cognitivo exigido para realizar tal tarefa é muito baixo.

Era desejável que tais sujeitos estivessem pelos menos no nível “leitura entre os dados” onde o sujeito é capaz de interpretar os dados e fazer uso de outras habilidades matemáticas como adicionar, subtrair, multiplicar e dividir os dados a fim de comparar quantidades, combinar e integrar as informações contidas no gráfico, uma vez que este é o nível de compreensão mais fácil de ser encontrado nas diversas atividades escolares e nos diferentes tipos de testes.

Dessa forma, sentimo-nos inclinadas a considerar que o trabalho estatístico efetuado no nível superior, privilegiando questões que extraem dos gráficos somente os dados explícitos em detrimento dos outros, não garante que os sujeitos dominem completamente a leitura e interpretação de gráficos no nível “leitura dos dados”, muito menos nos níveis “leitura entre os dados” e “leitura além dos dados”.

- **Qual é o nível de representação tabular, que o GFLM, o GFP e o GFADM apresentam?**

Apurados os resultados dos desempenhos dos sujeitos no teste, e, baseados nos preceitos da teoria de Wainer (1992) que definiu três níveis de leitura para tabelas, podemos responder que:

- Os sujeitos do grupo dos futuros licenciados em Matemática (GFLM) se encontram enquadrados no “nível intermediário”, no entanto de forma insatisfatória;
- Os sujeitos do grupo dos futuros Pedagogos (GFP) se enquadram no “nível básico”, porém de modo sofrível;
- Os sujeitos do grupo dos futuros Administradores (GFADM) se enquadram no “nível básico”, de maneira insatisfatória.

Para Wainer (1992) o “nível básico” é aquele em que as questões somente extraem da tabela os dados que estão explícitos e “nível intermediário” é aquele em que as questões exigem a interpolação ou a percepção da relação existente entre os dados de uma tabela.

Estes resultados corroboram com as ideias de Flores e Moretti (2006) quando argumentam que a simplicidade de uma tabela é somente aparente e que a organização representacional e visual e os processos cognitivos que ela desencadeia, podem interferir na leitura e análise dos dados nela contido.

Nesse sentido, nos sentimos confortáveis em dizer que as tabelas, no ensino destes sujeitos, estão sendo relegados a um segundo plano. Intuímos ainda, que provavelmente, tanto as escolas como as instituições de ensino superior trabalham aquelas tabelas consideradas por Wainer (1992) como “pobres”, por explorarem apenas o nível básico.

- **Estes três grupos de alunos, GFLM, GFP ou GFADM sabem converter dados estatísticos apresentados na forma de tabelas em gráficos e vice-versa, segundo Duval (2005)?**

Antes mesmo de responder tal questão, queremos ressaltar que, segundo Flores e Moretti (2006) a mudança de um registro em outro não é tão simples como possa parecer e que o ensino da Estatística privilegia muito mais a leitura e identificação de dados registrados em tabelas ou gráficos do que a própria construção destas representações.

Os resultados de nossas pesquisas corroboram com as ideias de Flores e Moretti (2006). Os três grupos participantes obtiveram um desempenho sofrível nas questões de conversão de tabelas para gráfico e vice-versa.

Ressaltamos que ao analisarmos as tabelas e gráficos construídos pelos sujeitos, consideramos como corretas apenas as respostas que continham todos os elementos necessários para tal construção, justamente por trabalharmos com o nível superior e julgarmos que são elementos essenciais para alunos de tal nível. No entanto, tivemos cerca de 40% dos sujeitos convertendo do registro gráfico para o registro tabular ou vice-versa, faltando algum dos elementos da construção.

Constatamos que os três grupos de alunos convertem do registro gráfico para o registro tabular ou do registro tabular para o registro gráfico de maneira insatisfatória.

- **Qual grupo tem melhor desempenho na conversão de dados estatísticos apresentados na forma de tabelas para gráficos e vice-versa?**

De acordo com os resultados apurados em nossa análise quantitativa, podemos apontar que todos os grupos obtiveram um desempenho sofrível na conversão de dados estatísticos apresentados na forma de tabelas para gráficos e vice-versa.

O grupo dos futuros licenciados em Matemática obteve uma taxa inferior a 6% nas questões que abordavam tanto conversão de gráficos para tabelas quanto o inverso. O grupo de futuros Pedagogos e Administradores obtiveram cada um uma taxa inferior a 3% em cada categoria.

- **Quais dos grupos se saem melhor em questões que envolvem cálculos de medidas de tendência central: o GFLM, o GFP ou o GFADM?**

Estamos considerando, dentro das medidas de tendência central os cálculos da Moda, Média e Mediana.

O grupo dos futuros licenciados em Matemática (GFLM) lidera as taxas de acertos nos três tipos de cálculos, mesmo não atingindo uma taxa considerada satisfatória. Nas questões que abordam o conceito de Moda, o GFLM obteve um desempenho satisfatório. No entanto, nas questões que abordavam tanto a Média quanto a Mediana, seu desempenho foi sofrível.

O grupo dos futuros Pedagogos (GFP) obteve desempenho insatisfatório nas questões que abordam o conceito de Moda e desempenho sofrível nos conceitos de Média e Mediana.

O grupo dos futuros Administradores obteve um desempenho satisfatório em questões que abordavam a Moda e insatisfatório nas questões que abordavam a Média e Mediana.

Nesse sentido reiteramos que o grupo GFLM se sai melhor em questões que envolvem cálculos de medidas de tendência central.

Diante das respostas aqui explicitadas e com base nas análises efetuadas, retomamos a questão principal de nossa pesquisa:

*QUAIS SÃO OS CONHECIMENTOS SOBRE LEITURA E INTERPRETAÇÃO DE GRÁFICOS E TABELAS QUE ALUNOS DOS CURSOS DE LICENCIATURA EM MATEMÁTICA, PEDAGOGIA E BACHARELADO EM ADMINISTRAÇÃO MOBILIZAM AO RESOLVER SITUAÇÕES APRESENTADAS EM FORMA DE PROBLEMAS?*

O grupo dos futuros licenciados em Matemática mobiliza os conhecimentos no nível “leitura dos dados” quando da leitura e interpretação de gráficos e no “nível intermediário” quando da leitura e interpretação de tabelas.

O grupo dos futuros Pedagogos mobiliza os conhecimentos no nível “leitura dos dados” quando da leitura e interpretação de gráficos e no “nível básico” quando da leitura e interpretação de tabelas.

O grupo dos futuros Administradores mobiliza os conhecimentos no nível “leitura dos dados” quando da leitura e interpretação de gráficos e no “nível básico” quando da leitura e interpretação de tabelas.

É importante ressaltar que os desempenhos do GFADM seguem muito próximo dos desempenhos do GFLM em todas as unidades de análise, salvo naquelas categorias que exigiam um custo cognitivo mais elevado como nos níveis “leitura além dos dados” e “avançado”, onde o GFLM se sai melhor.

Tendo ciência desses resultados, refutamos nossa hipótese de pesquisa de que os alunos de Administração apresentariam melhor desempenho frente às situações que envolvem a leitura e interpretação de gráficos e tabelas nos testes aplicados do que os alunos de Licenciatura em Matemática e Pedagogia.

De fato, esse grupo mostrou desempenho marcadamente melhor do que o grupo de Pedagogia e muito próximo do grupo de futuros licenciados em Matemática que, quase se equipara ao desempenho dos Administradores em algumas categorias. Porém o grupo que se mostrou com melhor desempenho foi o dos futuros licenciados em Matemática. E esse melhor desempenho está longe de ser considerado satisfatório, já que a média de acerto desse grupo ficou na casa dos 40%. E mais, o bom desempenho se resumiu à leitura e interpretação de gráfico (mais de 60%).

Assim, concluímos que os conhecimentos estatísticos dos três grupos mostram-se insuficientes, principalmente quando lhes são requeridos maiores esforços cognitivos para leitura e interpretação, em especial nas tabelas.

Contudo, chama atenção o pouco conhecimento apresentado pelos alunos de Pedagogia. Isto porque esses alunos, muito provavelmente, serão os responsáveis por introduzir os estudantes das séries iniciais do Ensino Fundamental no mundo da Estatística.

#### **5.4. Sugestão para futuras pesquisas**

Nossa pesquisa buscou responder algumas questões sobre o nível de Leitura e Interpretação de Gráficos e Tabelas que alunos do ensino superior apresentam, e, não tem a pretensão de encerrar o assunto. Ao longo de nosso trabalho, alguns

outros questionamentos surgiram e nos inspiraram a sugerir algumas ideias para novas pesquisas.

Dessa forma, destacaremos quatro sugestões:

1ª. Investigar quais são os conhecimentos sobre leitura e interpretação de gráficos e tabelas que os atuais professores de Matemática em exercício de sua profissão, mobilizam ao resolver situações apresentadas em forma de problemas. Possivelmente, grande parte dos professores de Matemática não estão preparados para inserir questões enquadradas nos níveis “leitura além dos dados” e “nível avançado” em suas aulas.

Nesse sentido sugerimos uma intervenção de ensino voltada para tais educadores, por serem eles os responsáveis pelo ensino da Estatística no ensino fundamental e médio. Sugerimos ainda, uma investigação com professores da rede estadual que atuem no ensino fundamental II e ensino médio, por acreditarmos que uma boa parcela dos professores de Matemática apresenta dificuldades com relação à Estatística.

Tal investigação partiria primeiramente, de um teste diagnóstico que detecte quais são os níveis de leitura e interpretação de gráficos segundo Curcio (1989) nos quais estes se enquadram. Na sequência, elaborar uma intervenção nas aulas de Estatística que propicie um (re) delineamento e (re) avaliação, com vistas na documentação das ações, das interações e dos efeitos desta interferência com os sujeitos da pesquisa.

2º. Desenvolver um estudo com professores de Matemática e Pedagogia das séries iniciais da rede estadual e ensino, sobre leitura e interpretação de gráficos e tabelas a fim de munir tais professores com subsídios para trabalhar com seus alunos gráficos e tabelas dentro dos três níveis de Curcio (1989) e de Wainer (1992).

Pautados nessas ideias sugerimos ainda, que seja implementada uma sequência didática para a abordagem da Estatística, mais especificadamente a leitura e interpretação de gráficos e tabelas. Tal implementação poderia acontecer em um número de encontros pré-determinado pelo grupo de professores dispostos a participar do estudo.

3º. Desenvolver pesquisa com teste-diagnóstico e intervenção de ensino pautada na análise exploratória dos dados. Esta pesquisa teria como sujeitos

professores de Matemática do ensino fundamental II e ensino médio da rede pública estadual.

Como nossa pesquisa aponta uma dificuldade dos sujeitos em resolver questões baseadas na teoria dos Registros de Representação Semiótica voltadas para a Estatística, pensamos que tal sugestão só agregaria conhecimentos aos professores participantes da pesquisa.

4º. Desenvolver um estudo voltado para o ensino da Estatística nos diferentes cursos superiores. Investigar se o ensino da Estatística em cursos de licenciaturas foca, ou pelo menos determina, um momento para discussão de como está ou como deveria ser o ensino da Estatística nas salas de aulas do ensino fundamental e médio.

Também acreditamos que seja necessário investigar se a Estatística ensinada em outros cursos, diferentes de licenciatura e que farão uso desta Estatística, está efetivamente munindo esses futuros profissionais com ferramentas estatísticas suficientes para que estes realizem a leitura adequada e a interpretação de gráficos e tabelas em seu cotidiano profissional.

Acreditamos que estudo como esses poderão trazer muita contribuição para o entendimento e o avanço do processo da aprendizagem da Estatística, além de certamente trazer melhoria no ensino de tal disciplina.

## REFERÊNCIAS

---

- ABRANTES, Paulo. SERRAZINA, Lurdes. OLIVEIRA, Izolina. A Matemática na educação Básica. Lisboa: Ministério da Educação. 1999
- ALMOULOU, Saddo Ag. Fundamentos da didática da matemática. Curitiba: Ed. UFPR, 2007.
- ARAÚJO, Elizangela Gonçalves de. Flores, Cláudia Regina. (2005). *O tratamento da Informação nas Séries Iniciais*: uma proposta de formação de professores para o ensino dos gráficos e tabelas. Disponível em: [http://www.sbem.com.br/files/ix\\_enem/Poster/Trabalhos/PO02436001944T.doc](http://www.sbem.com.br/files/ix_enem/Poster/Trabalhos/PO02436001944T.doc). Acesso em 08 fev 2009.
- BATANERO, C. ESTEPA, E. GODINO, J. D. Análisis exploratório de datos: sus posibilidades en la enseñanza secundaria,. *Suma*, n° 9, p.25-31, 1991. Disponível em: <<http://www.ugr.es/~batanero>>. Acesso em 26 mar. 2008.
- BIAJONE, Jefferson. Trabalho de Projetos: Possibilidades e Desafios na Formação Estatística do Pedagogo. São Paulo, 2006. Dissertação (Mestrado em Educação Matemática). Universidade Estadual de Campinas, Faculdade de Educação.
- BRASIL. Ministério da Educação. Secretaria de Educação Fundamental. Parâmetros Curriculares Nacionais: matemática. Brasília: Ministério da Educação/Secretaria de Educação Fundamental, 1998.
- BRASIL. Ministério da Educação. Secretaria da Educação Média e Tecnologia. Parâmetros Curriculares Nacionais. Brasília: Ministério da Educação, 1999.
- BRASIL. Ministério da Educação. Secretaria da Educação Média e Tecnologia. Orientações Complementares aos Parâmetros Curriculares Nacionais. Brasília: Ministério da Educação, 2002.
- CARVALHO, Carolina. Literacia Estatística. In: I Seminário de Ensino de Matemática – 14ª Conferência, Campinas: São Paulo, 2003. p. 22-25.
- CAZORLA, Irene Maurício. A relação entre a habilidade viso-pictória e o domínio de conceitos estatísticos na leitura de gráficos. São Paulo, 2002. Tese (Doutorado). Universidade Estadual de Campinas, Faculdade de Educação.
- CARZOLA, Irene Maurício. CASTRO, Franciana Carneiro de. O Papel da Estatística na Leitura do Mundo: O Letramento Estatístico. *Publ. UEPG Ci. Hum., Ci. Soc. Apl., Letras e Artes*, Ponta Grossa, 16 (1) 45-53, jun. 2008.
- CARZOLA, Irene Mauricio. SANTANA, Eurivalda Ribeiro dos Santos. Tratamento da informação para o ensino fundamental e médio. – Itabuna: Via Literarum, 2006.

CURCIO, F.R. Developing graph comprehension. Virginia: National Council of Teachers of Mathematics. ISBN 0-87353-277-5. pp5-6, 1989.

DAMM, R. F. Registros de Representação. In: MACHADO, S. D. A. et al. *Educação Matemática: uma introdução*. São Paulo: EDUC, 1999, p.135-153.

DUVAL, R. (2003) Presentation. Spirale Revue de Recherches em Éducation, n.32. p. 3-6. Disponível em: <http://arred.free.fr/spirale/IMG/pdf/32-PreSentation-Spirale-32.pdf>. Acesso em: 08 fev. 2009.

\_\_\_\_\_. Registros de representações semióticas e funcionamento cognitivo da compreensão em Matemática. In: MACHADO, S. D. A. (org.). *Aprendizagem em Matemática: Registros de Representação Semiótica*. Campinas, SP: Papyrus, 2005, p.11-33.

Educación Estadística em La Matemática Escolar: retos para La Enseñanza y La Formación Del Profesor (Documento de discusión). UNION Revista Iberoamericana de Educación Matemática. n.8, p.63-75, dez. 2006. Disponível em: <http://www.fisem.org/paginas/union/revista.php?id=22#indice>. Acesso em: 26 mar 2008,

FLORES, Cláudia R. MORETTI, Mércles T. (2006) *O funcionamento cognitivo e semiótico das representações gráficas*: ponto de análise para a aprendizagem matemática. Disponível em: [http://www.redemat.mtm.ufsc.br/reremat/republic\\_07\\_artigo.PDF](http://www.redemat.mtm.ufsc.br/reremat/republic_07_artigo.PDF). Acesso em: 08 fev. 2009.

FREITAS, Henrique et al. O método de pesquisa survey. Revista de administração. São Paulo, v. 35, n. 3, p. 105-112, jul./set. 2000. Disponível em: <http://www.rausp.usp.br/>. Acesso em: 28 jan. 2009.

FIORENTINI, Dario. LORENZATO, Sergio. *Investigação em educação matemática: percursos teóricos e metodológicos*. São Paulo: Autores Associados, 2006.

GAL, Iddo. *Adults' Statistical literacy*: Meanings, components, responsibilities. International Statistical Review. Israel, 70(1), p.1-25, 2002.

GRÁCIO, Cláudia Cabrini.; OLIVEIRA, Ely Francina Tannuri de; OLIVEIRA, Sônia Marina Tavares de. (s.d). *A disciplina Estatística nos cursos de Graduação da UNESP do Campus de Marília: Avaliação de um Projeto de Ensino*. Disponível em: <http://www.posgraduacao.ufla.br/gauss/congresso/47rbras/e1.pdf> Acesso em 15 mar. 2009.

GUEDES, Terezinha Aparecida. et al. *Projeto de Ensino: Aprender fazendo Estatística*. Disponível em: [http://www.des.uem.br/downloads/arquivos/Estatistica\\_Descritiva.pdf](http://www.des.uem.br/downloads/arquivos/Estatistica_Descritiva.pdf) Acesso em 08 fev 2009.

LEVIN, Jack; Fox, James Alan. *Estatística para Ciências Humanas*. 9. ed. Tradução de Alfredo Alves de Farias. São Paulo: Pearson Education do Brasil, 2007. 497 p. Título original: Elementary statistics in social research.

MARCONI, Marina de Andrade. LAKATOS, Eva Maria. *Metodologia do Trabalho Científico: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos*. 7. ed. São Paulo: Atlas, 2008.

MÓDOLO, Cristiane Machado; GOUVEIA Jr., Amauri. Estudo quantitativo dos infográficos publicados na revista Superinteressante nos anos de 1987 a 2005. In: XXX Congresso Brasileira de Ciências da Comunicação, Santos: São Paulo, 2007.

NOVAES, Diva Valério; Coutinho, Cileda Q. S. *Estatística para Educação Profissional*. São Paulo: Editora RBD, 2008.

PEREIRA, Sérgio Alves. Um estudo a respeito do professor de Matemática e a Implementação de uma sequência didática para a abordagem da Estatística no Ensino Médio. São Paulo, 2007. Dissertação (Mestrado em Educação Matemática). Pontifícia Universidade Católica de São Paulo.

RIBEIRO, José Odair. *Leitura e Interpretação de Gráficos e Tabelas: um estudo exploratório com professores*. São Paulo, 2007. Dissertação (Mestrado em Educação Matemática). Pontifícia Universidade Católica de São Paulo.

SEVERINO, Antonio Joaquim. *Metodologia do Trabalho Científico*. 23ª ed.; São Paulo: Cortez Editora, 2008.

SILVA, Cláudia Borim da. *Pensamento estatístico e raciocínio sobre variação: um estudo com professores de matemática*. São Paulo, 2007. Tese (Doutorado em Educação Matemática). Pontifícia Universidade Católica de São Paulo.

SILVA, Cláudia Borim da. et al. Atitudes em relação à estatística e à matemática. *Psico-USF*, Campinas, v.7, n.2, p. 219-228, jul./dez. 2002.

SILVA, Edgard Dias da. *Os conceitos elementares de Estatística a partir do Homem Vitruviano: uma experiência de ensino em ambiente computacional*. São Paulo, 2008. Dissertação (Mestrado em Educação Matemática). Pontifícia Universidade Católica de São Paulo.

SOUZA, Paulo Nathanael Pereira de. SILVA, Eurides Brito da. *Como entender e Aplicar a Nova LDB (Lei nº 9.394/96)*. São Paulo: Pioneira Thomson Learning, 2001.

VASCONCELOS, Paulo Ramos. *Leitura e Interpretação de gráficos e tabelas: Estudo exploratório com alunos da 8ª série do ensino fundamental*. São Paulo, 2007. Dissertação (Mestrado em Educação Matemática). Pontifícia Universidade Católica de São Paulo.

VASQUES, Ricardo Sergio Braga. *Mobilização de conceitos estatísticos: um estudo diagnóstico desses conceitos envolvendo variabilidade com alunos do Ensino Médio*. São Paulo, 2007. Dissertação (Mestrado em Educação Matemática). Pontifícia Universidade Católica de São Paulo.

VENDRAMINI, Claudette Maria Medeiros. BRITO, Márcia Regina Ferreira de. Relações entre atitude, conceito e utilidade da estatística. *Psicol. esc. educ.* São Paulo, v. 5, n. 1, p. 59-73, jun.2001. Disponível em: [http://pepsic.bvs-psi.org.br/scielo.php?script=sci\\_arttext&pid=S1413-85572001000100007&lng=PT&nr](http://pepsic.bvs-psi.org.br/scielo.php?script=sci_arttext&pid=S1413-85572001000100007&lng=PT&nr)

m=isso. Acesso em: 28 jan. 2009.

WADA, Ronaldo Seichi. *Estatística e Ensino: um estudo sobre representações de professores do 3º grau*. São Paulo, 1996. Tese (Doutorado). Universidade Estadual de Campinas, Faculdade de Educação.

WAINER, Howard. Understanding Graphs and Tables. *Educational Research*, v. 21, n. 1, p. 14-23, jan/fev. 1992.

WALLMAN, Katherine K. (1993). Enhancing Statistical literacy: enriching our society. *Journal of the American Statistical Association*, 88 (421), 1-8.

**APÊNCIDE A**  
**O INSTRUMENTO DIAGNÓSTICO**

---

**Questionário sobre Tratamento da Informação**

Caro(a) Aluno(a)

O presente questionário é um instrumento para coleta de dados de uma pesquisa em Educação Matemática ainda em andamento, para obtenção de título de Mestre em Educação Matemática do Programa de Estudos Pós-graduados da PUC-SP, sob orientação da Prof<sup>a</sup> Dr<sup>a</sup> Sandra Magina. Sua participação é importante para nós e nos comprometemos mantê-lo(a) informado(a) sobre as análises que faremos a partir do material levantado. Agradecemos imensamente sua colaboração.

**1. PERFIL**

1.1 Gênero:

Masculino                       Feminino

1.2. Qual a sua faixa etária?

18 a 24 anos                       37 a 42 anos  
 25 a 30 anos                       43 ou mais  
 31 a 36 anos

1.3. Você lembra se estudou Estatística na escola **antes da graduação**?  Sim               Não

1.4. Se sim, em que ano? \_\_\_\_\_

1.5. O que lembra de ter estudado sobre Estatística? \_\_\_\_\_

1.6. Você está cursando

Licenciatura em Matemática  
 Pedagogia  
 Administração de Empresas

1.6.1. Quantos anos ou semestres tem o seu curso? \_\_\_\_\_

1.7. Qual o ano ou semestre que você está cursando? \_\_\_\_\_

1.8. Durante a sua graduação você já estudou Estatística?  Sim       Não

1.9. Em que ano ou semestre? \_\_\_\_\_

1.10. Quais os conteúdos Estatísticos já estudados na graduação? \_\_\_\_\_

1.11. Destes conteúdos estudados, quais deles você considera mais importante? \_\_\_\_\_

\_\_\_\_\_ Por quê? \_\_\_\_\_

1.12. Você está lecionando?  Sim       Não

1.13. Em que nível de escolaridade você leciona?

Não leciono  
 Ensino Fundamental I               Matemática  Outra disciplina  
 Ensino Fundamental II               Matemática  Outra disciplina  
 Ensino Médio                       Matemática  Outra disciplina  
 Superior                               Matemática  Outro curso

1.14. Tempo de Magistério: \_\_\_\_\_  Não leciono

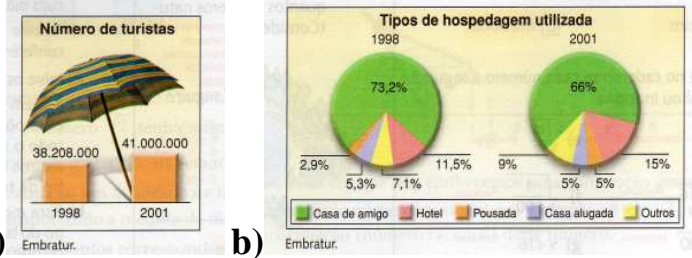
1.15. Você considera o Ensino de Estatística:

muito importante  
 importante  
 pouco importante  
 não é importante  
 não sei avaliar

1.16. Dê três razões de sua escolha na questão 1.15 \_\_\_\_\_

---

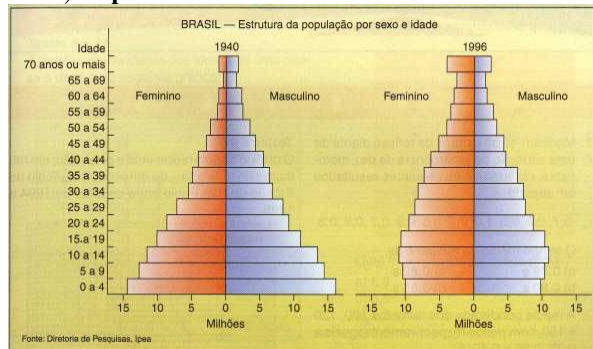
**2. Observe os gráficos abaixo de uma pesquisa sobre o turismo doméstico e responda:**



Fonte: Projeto Araribá: Matemática/obra coletiva. São Paulo: Moderna, 2006.

- 2.1. Para saber a quantidade de turistas que se hospedou em hotel no ano de 2001, precisamos ler as informações de qual desses gráficos?**
- ( ) Somente o gráfico (a)
  - ( ) Somente o gráfico (b)
  - ( ) Os dois, (a) e (b)
  - ( ) Nenhum dos dois
- 2.2. Qual a informação que o gráfico (a) traz?**
- ( ) A quantidade de turistas que realizaram viagens no Brasil no ano de 1998
  - ( ) A quantidade de turistas que realizaram viagens no Brasil no ano de 2001
  - ( ) A quantidade de turistas que realizaram viagens no Brasil no ano de 1998 e 2001
  - ( ) A quantidade de turistas que realizaram viagens no Brasil no ano de 1998 a 2001
- 2.3. O gráfico (b) traz que tipo de informação?**
- ( ) O percentual de turistas de acordo com os tipos de transporte utilizado.
  - ( ) O número de turistas de acordo com o tipo de hospedagem utilizada.
  - ( ) O percentual de turistas de acordo com o tipo de hospedagem utilizada nos anos de 1998 e 2001.
  - ( ) O percentual de turistas de acordo com o tipo de hospedagem utilizada no ano de 1998.

**3. Com base nos gráficos abaixo, responda:**



Fonte: Iezzi, G. Dolce, O. Machado, A. - Matemática e Realidade: 7ª série - 4 ed. Reform. - São Paulo: Atual, 2000.

- d) Em 1940, na faixa etária de 0 a 4 anos, havia mais homens ou mais mulheres no Brasil? \_\_\_\_\_ E em 1996? \_\_\_\_\_
- e) Em que faixa estava a idade modal masculina em 1996? \_\_\_\_\_
- f) Qual a faixa etária que tem menos população em 1940? \_\_\_\_\_

**4. Na tabela abaixo estão computadas as opiniões fictícias de 60 pessoas sobre um filme que acabava de estreiar na cidade de Argolina.**

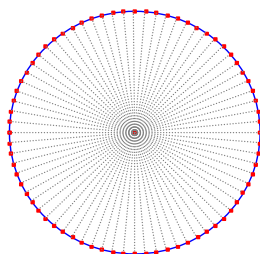
OPINIÃO	Nº DE PESSOAS
Excelente	9
Ótimo	15
Bom	18
Regular	12
Ruim	3
Péssimo	3
TOTAL	60

Fonte: Iezzi, G. Dolce, O. Machado, A. - Matemática e Realidade: 6ª série - São Paulo: Atual, 2000.

<p>4.1. É possível calcular a média dos dados acima?                  ( ) Sim ( ) Não                  Por quê?</p>	<p>4.2. Indique a opinião mais freqüente (moda).</p>
---	--

4.4 Represente os dados da tabela em um gráfico

2.4. Calcule as porcentagens relativas às opiniões e represente-as num gráfico de setores. Para esta construção utilize a circunferência abaixo que já está dividida de 5° em 5°.



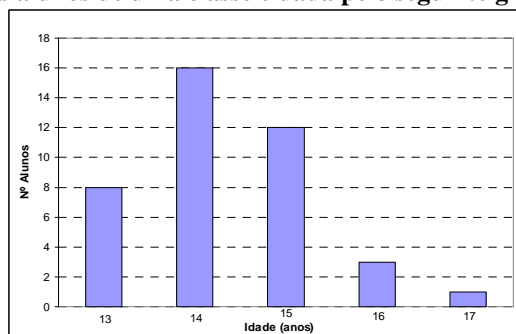
5. Uma prova contendo cinco questões, cada uma valendo 2,0 pontos, foi aplicada numa classe de 40 alunos. Na tabela está o número de alunos que acertou cada questão:

Questão	Número de alunos que acertaram
1	12
2	40
3	24
4	32
5	36

Fonte: lezzi, G. Dolce, O. Machado, A. - Matemática e Realidade. São Paulo:Atual, 2000.

5.1. Onde houve maior variação do número de acertos?	5.2. Qual é a moda da questão mais respondida?
--	--

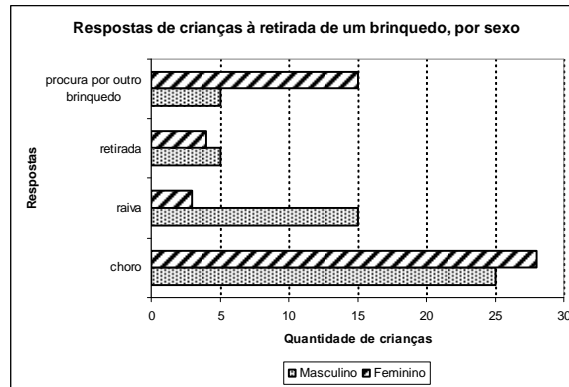
6. A distribuição de idades dos alunos de uma classe é dada pelo seguinte gráfico:



Fonte: lezzi, G. Dolce, O. Machado, A. - Matemática e Realidade. São Paulo:Atual, 2000.

6.1. Qual é a idade média dos alunos?	6.2. A idade modal é:
	6.3. Qual é a idade que representa a mediana?

7. Observe o gráfico abaixo:



Fonte: Adaptado de Levin, Jack. Fox, J.A; Estatística para ciência Humanas. São Paulo: Pearson Prentice Hall, 2004.

Agora com base na leitura dos dados contidos no gráfico de barras acima, construa uma tabela. Para facilitar a construção tem-se abaixo uma grade. Utilize o quanto for necessário e risque o restante.


**8. Analise as informações da tabela:**

Uso do cinto de segurança	Sexo do entrevistado		Total
	Masculino	Feminino	
Sempre	144	355	499
Na maioria das vezes	66	110	176
Algumas vezes	58	66	124
Raramente	39	44	83
Nunca	60	55	115
Total	367	630	997

Fonte: Adaptado de Levin, Jack. Fox, J.A; Estatística para ciência Humanas. São Paulo: Pearson Prentice Hall, 2004.

Agora com base na leitura da tabela acima construa um gráfico, por sexo, do entrevistado:

**9. (ENEM/2001) A tabela apresenta a taxa de desemprego dos jovens entre 15 e 24 anos estratificada com base em diferentes categorias.**






Região	Homens	Mulheres
Norte	15,3	23,8
Nordeste	10,7	18,8
Centro-oeste	13,3	20,6
Sul	11,6	19,4
Sudeste	16,9	25,7
Grau de Instrução		
Menos de 1 ano	7,4	16,1
De 1 a 3 anos	8,9	16,4
De 4 a 7 anos	15,1	22,8
De 8 a 10 anos	17,8	27,8
De 11 a 14 anos	12,6	19,6
Mais de 15 anos	11,0	7,3

Fonte: PNAD/IBGE, 1998.

Considerando **apenas** os dados acima e analisando as características de candidatos a emprego, é possível concluir que teriam **menor** chance de consegui-lo,

- (A) Mulheres, concluintes do ensino médio, moradoras da cidade de São Paulo.
- (B) Mulheres, concluintes de curso superior, moradoras da cidade do Rio de Janeiro.
- (C) Homens, com curso de pós-graduação, moradores de Manaus.
- (D) Homens, com dois anos do ensino fundamental, moradores de Recife.
- (E) Mulheres, com ensino médio incompleto, moradoras de Belo Horizonte.

**10. Na sala de aula de 4ª série J, a professora de Matemática solicitou que cinco alunos se colocassem em pé na frente da sala, medindo com uma fita métrica a estatura de cada um e anotando na lousa. Veja a estatura dos cinco alunos.**

	Luiz 152 cm		Ana 148 cm		João 155 cm		
		Bia 145 cm				Caio 150 cm	

Adaptado de Carzola e Santana, Tratamento da Informação para o Ensino Fundamental e Médio. Bahia, Via Litterarum 2006.

a) Com base na leitura dos dados acima, construa uma tabela. Para facilitar a construção tem-se abaixo uma grade. Utilize o quanto for necessário e risque o restante.


b) Agora de acordo com a tabela que você construiu, calcule a altura mediana dos meninos da 4ª série J.

**MEDIDAS DE TENDÊNCIA CENTRAL:  
A ORGANIZAÇÃO E APRESENTAÇÃO DOS DADOS**

---

A Estatística é uma área do saber que se encarrega, também, da análise de dados e parece-nos que cada vez mais desempenha um papel relevante em praticamente todas as áreas do conhecimento. A Estatística é uma ferramenta amplamente utilizada, por exemplo, em pesquisas, e que essa pesquisa a tem como tema, então devemos conhecer a definição de Estatística.

Isto posto, iremos aceitar como definição de Estatística, em meio a tantas, a de Barnett: “*Estatística é o estudo de como a informação deveria ser empregada para reflexão e ação em uma situação prática envolvendo incerteza*” (BARNETT, 1973 apud NOVAES e COUTINHO, 2008, p.5). Tal opção encontra apoio na ideias de Novaes e Coutinho (2008) quando argumentam que:

Vamos assim assumir a Estatística como sendo a ciência que considera o número em contexto, ou seja, vamos considerar os resultados obtidos na resolução de um problema como sendo relevantes se e somente se analisados no contexto desse problema no qual foram encontrados. (NOVAES; COUTINHO, 2008, p.5)

Nesse sentido, corroborando das idéias de Novaes & Coutinho (2008) e analisando os resultados no contexto, temos que a **Estatística Descritiva** é uma ferramenta da qual fazemos uso desde o planejamento da pesquisa e coleta de dados até a organização destes e suas diversas representações, em outras palavras a Estatística Descritiva descreve um conjunto de dados.

Uma vez descrito o conjunto de dados, precisamos partir para as inferências - estamos no campo da **Estatística Inferencial**. É importante ressaltar que fazer inferência significa “*tomar decisões com base em dados coletados de apenas uma parte ou amostra do grupo maior que deseja estudar*” (LEVIN; FOX, 2007, p.15)

Concordamos com Novaes e Coutinho (2008) ao dizer que:

Quando se quer inferir fatos acerca da população, a partir do estudo de uma parte significativa dessa população (amostra), estamos no campo da Inferência Estatística. Considerando ainda, os casos em que uma análise é realizada com os dados de uma amostra, necessitamos de algumas noções de inferência para estender os resultados para a população de onde foram extraídos os dados. Dessa forma pode-se extrair o máximo de informações possíveis dos dados com as ferramentas da Estatística Descritiva e noções de Inferencial. (NOVAES; COUTINHO, 2008, p.5)

Nessa linha de raciocínio, encontramos a Estatística dividida em duas partes: Estatística Descritiva e Estatística Inferencial, tendo como elo entre as duas partes a Teoria de Probabilidades. Quando partimos para a utilização de todas as partes da Estatística na busca da resolução de problemas, estamos enveredando pelo caminho de uma nova filosofia chamada **análise exploratória de dados**.

Encontramos a definição dessa nova filosofia, a análise exploratória de dados, nas palavras de Batanero, Estepa e Godino (1991), a saber:

Esta filosofia consiste no estudo dos dados a partir de todas as perspectivas e com todas as ferramentas possíveis, incluindo as já existentes. O propósito é extrair toda a informação possível, gerar novas hipóteses no sentido de construir conjecturas sobre as observações que dispomos. (BATANERO, C; ESTEPA, A; GODINO, J D, 1991, p.2)

Acreditamos ser este o caminho para uma construção sólida da Educação Estatística, e para tanto, devemos conhecer alguns conceitos elementares estatísticos, tais como: variável, população, amostra, amostragem e distribuição de frequências.

Para entendermos o significado do termo **variável**, tomemos como base a pesquisa social. Entendemos que muitos de nós, talvez de uma maneira rudimentar, temos dentro de si um pouco de pesquisador social. Praticamente todos os dias, arquitetamos planos para o nosso futuro, fazemos previsões a fim de vivenciar novas situações ou experiências. Quando estes planos dão certo só confirmam ou apóiam nossas ideias iniciais, porém quando não somos tão felizes, devemos enfrentar implicações nem sempre muito agradáveis.

Assim como em nosso cotidiano, o pesquisador social procura explicar o comportamento humano por meio de previsões construtivas de maneira muito mais precisa e estruturada teórico-metodologicamente. No desenvolver desse processo o pesquisador examina características do comportamento humano que diferem ou variam de um indivíduo para o outro – são as **variáveis**. Por exemplo, a cor do cabelo, a idade, estatura, comportamento, etc., ou características que variam de um

instante para outro no decorrer do tempo como, por exemplo, desemprego, taxa de mortalidade infantil, população, etc. Cf. LEVIN, J; FOX, J. A. Estatística para Ciências Humanas, p.1.

Partindo do pressuposto que uma variável pode representar características que diferem de indivíduo para indivíduo, então podemos questionar quais são as diferenças existentes entre esses tipos de variáveis?

Para respondermos esta questão, vamos supor que um pesquisador canadense queira analisar a atitude de 10 estudantes universitários norte-americanos em relação aos latinos. A variável “atitude em relação aos latinos” denota uma característica de comportamento daquele grupo pesquisado, portanto estamos falando de uma **VARIÁVEL QUALITATIVA**. Em outras palavras, “a *variável qualitativa revela um certo tipo de características relacionadas ao grupo pesquisado, por exemplo, cor dos olhos, sexo, grau de satisfação com os serviços prestados e outras*” (NOVAES; COUTINHO, 2008, p.16).

Quando estas variáveis qualitativas são passíveis de mensuração e nos permitem associar um número a elas, seguindo um conjunto de regras, estas se dividem em: **NÍVEL NOMINAL** e **NÍVEL ORDINAL** de mensuração.

Assim temos que “o *nível nominal de mensuração consiste em nomear ou rotular – isto é, dispor os casos em categorias e contar sua frequência de ocorrência.*” (LEVIN; FOX, 2007, p. 9).

As **variáveis qualitativas nominais** devem ser enquadradas em apenas uma categoria, assim retomando nosso exemplo, a atitude do entrevistado ou é preconceituosa ou tolerante, elas não devem sobrepor-se. Note que também os dados nominais, por vezes são rotulados por números (1= preconceituosa, 2= tolerante), mas com o objetivo de agrupar os casos em categorias diferentes e de maneira nenhuma quantificar as atitudes. Observe na figura 66 que apesar de utilizarmos os rótulo 1 e 2, eles não significam a magnitude da atitude:

TABELA 1.1 Atitudes de 10 estudantes universitários norte-americanos em relação aos latinos: dados nominais	
Atitude em relação aos latinos	Frequência
1 = preconceituosa	5
2 = tolerante	5
Total	10

Figura B.67: Exemplo de variáveis qualitativas nominais, retirada de LEVIN, J; FOX, J.A. 2007, p. 9.

Já as **variáveis qualitativas ordinais** são definidas para além do nível de mensuração, quando procuramos “ordenar seus casos em termos do grau em que possuem determinada característica” (LEVIN; FOX, 2007, p.10).

Quando as variáveis são mensuradas como ordinal, esta não indica a magnitude de diferenças entre números. Observe a figura 67.

TABELA 1.2 Atitudes de 10 estudantes universitários norte-americanos em relação aos latinos: dados ordinais	
Aluno	Posto
Joyce	1 = mais preconceituoso
Paul	2 = segundo
Cathy	3 = terceiro
Mike	4 = quarto
Judy	5 = quinto
Joe	6 = sexto
Kelly	7 = sétimo
Ernie	8 = oitavo
Linda	9 = nono
Ben	10 = menos preconceituoso

Figura B.68: Exemplo de variáveis qualitativas ordinais, retirada de LEVIN, J; FOX, J.A. 2007, p. 11.

Somos capazes de dizer o quão é mais preconceituosa a atitude de Joyce em relação à de Mike? Ou, quão é menos preconceituoso Ben de Linda ou Kelly? A resposta é não. Os intervalos entre os postos não significa “a quantidade de atitude preconceituosa” que cada um dos entrevistados tem, simplesmente pelo fato de que não “quantificamos” preconceito. O nível ordinal de mensuração proporciona apenas informação sobre a ordenação das categorias, desse modo, “*não podemos atribuir escores a casos localizados em pontos ao longo da escala*” (LEVIN; FOX, 2007, p.11).

Encontramos também, o **nível intervalar** de mensuração que além de informar a ordenação das categorias também indica a distância entre cada uma. Podemos perceber que algumas variáveis são naturalmente intervalares, como por exemplo, o número de irmãos ou a quantidade de horas que gastamos assistindo a televisão, enquanto outras apresentam nível intervalar estabelecido pela pessoa que está tratando os dados. Observe a figura 68:

<b>TABELA 1.3 Atitudes de 10 estudantes universitários norte-americanos em relação aos latinos: dados intervalares</b>	
<b>Aluno</b>	<b>Escore<sup>a</sup></b>
Joyce	98
Paul	96
Cathy	95
Mike	94
Judy	22
Joe	21
Kelly	20
Ernie	15
Linda	11
Ben	6

<sup>a</sup> Escores mais altos indicam maior preconceito em relação aos latinos.

Figura B.69: Exemplo de nível intervalar, retirada de LEVIN, J; FOX, J.A. 2007, p. 12.

Ainda com relação ao nosso exemplo, imaginem que o pesquisador canadense tenha formulado uma bateria de perguntas sobre latinos e que cada resposta obtida tivesse a medida intervalar de preconceito de 0 a 100, onde 100 é o preconceito máximo.

Nesse sentido, o pesquisador ordenou os estudantes em termos de seus preconceitos levantados nas respostas à pesquisa e indicou somente as distâncias que separam uns dos outros. Analisando a figura 5, podemos inferir que Joyce, Paul, Cathy e Mike são os mais preconceituosos, enquanto Ben, Linda, Ernie e Kelly são os menos preconceituosos da turma porque receberam os escores mais baixos. Observando o grupo dos menos preconceituosos, podemos dizer também, que entre Ben e Linda existe uma pequena diferença, em outras palavras, Linda é ligeiramente mais preconceituosa que Ben.

Além das variáveis qualitativas, temos as **VARIÁVEIS QUANTITATIVAS** que assumem valores numéricos e são definidas em um intervalo real, como por exemplo, o número de alunos que realizaram matrículas em escolas particulares,

número de filhos por família, números homens que freqüentam academias de ginástica, etc. Estas variáveis se dividem em **DISCRETA** ou **CONTÍNUA**.

A **variável quantitativa** é **discreta** quando “o conjunto de valores que ela assume é oriundo de uma contagem, ou seja, entre dois valores consecutivos da variável não podemos inserir nenhum outro valor” (NOVAES; COUTINHO, 2008, p.16). Vamos exemplificar: o número de filhos é uma variável quantitativa discreta, por que conseguimos fazer uma contagem e não há a possibilidade de um casal ter 2,7 filhos; isto significa que neste caso eles só poderão ter 2 ou 3 filhos.

Em contrapartida uma **variável quantitativa** se diz **contínua** quando pode assumir qualquer valor dentro de um intervalo real. É o caso, por exemplo, de queremos pesquisar o peso de todas as pessoas da cidade de “Luacheia” compreendido entre 67 Kg. e 68 Kg.; sempre há a possibilidade de incluirmos um novo valor.

Porém, no mundo real, fica muito difícil medir com precisão uma variável contínua. Nesse caso, na prática, podemos pensar em todas as medidas como sendo discretas. O problema se constitui na representação gráfica de variáveis contínuas que se diferenciam da representação gráfica das discretas, sem contar que algumas variáveis discretas também podem receber o tratamento de contínuas “devido à amplitude total determinada pelos valores observados” (NOVAES; COUTINHO, 2008, p.17).

Como forma de sintetizar estas considerações, observe a figura 6 e note como podemos categorizar as variáveis:

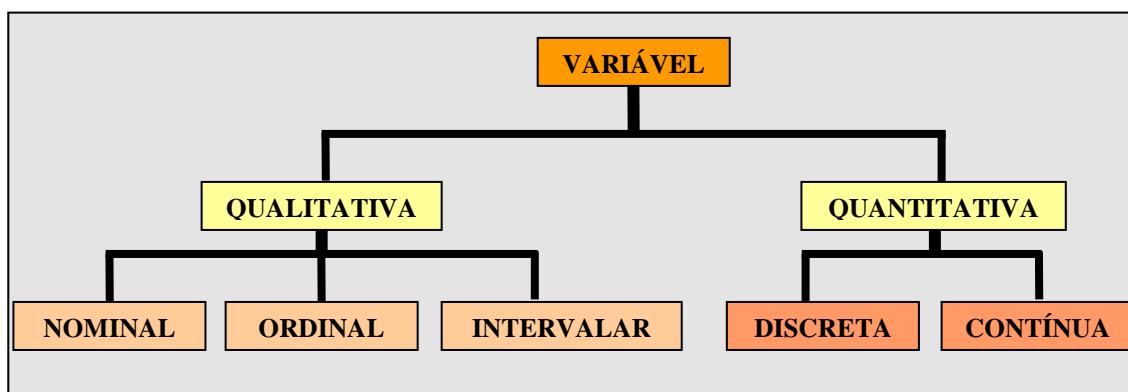


Figura B.70: Esquema de classificação de variáveis.

Após uma breve discussão sobre variáveis, vamos recuperar nosso exemplo no qual o pesquisador canadense estava interessado em saber qual a atitude de 10 estudantes universitários norte-americanos em relação aos latinos. Imaginemos que

antes de chegar a este número de estudantes, dez, o pesquisador estava interessado em saber a atitude de todos os estudantes universitários norte-americanos. Então, todos os estudantes passariam a ser um conjunto denominado **POPULAÇÃO**, pois é um conjunto definido e que contém a totalidade dos elementos que se quer estudar. Outros exemplos de população seriam a produção diária de queijo de uma indústria de laticínios, a venda total de veículos das indústrias automobilísticas localizadas na região do ABC Paulista no mês de Janeiro de 1998 ou o desempenho dos alunos de 5ª série da “Escola Ratatui” em cálculo de áreas e perímetro.

Retomando o exemplo do pesquisador canadense e sua pretensão de pesquisar todos os estudantes universitários norte-americano; podemos classificar essa tarefa como árdua e que provavelmente gastaria muito tempo, fora o fato de ter um custo elevado. Para esses casos, pesquisamos uma **AMOSTRA**. Em outras palavras, amostra é uma parte da população, um subconjunto finito daqueles elementos que se quer estudar. Hipoteticamente, vamos admitir que nosso pesquisador tenha escolhido como amostra para sua pesquisa, a “Universidade Alfa”, com cinco mil alunos matriculados– continuamos com um número grande.

Ainda no campo da suposição, vamos imaginar que nosso pesquisador tenha percebido que não seria viável pesquisar a amostra toda e resolveu fazer uma **AMOSTRAGEM**. Para Novaes e Coutinho (2008) amostragem consiste no:

*“... processo de fixar critérios para a composição de uma amostra que tenha a representatividade necessária no estudo em questão (...) que todos os elementos da população tenham a mesma chance de serem escolhidos...”*  
(NOVAES; COUTINHO, 2008, p. 18)

Então, podemos perceber que nosso pesquisador lançou mão de algum tipo de amostragem para chegar ao total de dez alunos pesquisados. Existem diversos tipos de amostragens e as probabilísticas mais utilizadas são a aleatória simples e a estratificada. Para esta pesquisa, não discutiremos os tipos de amostragens por acreditarmos se tratar de um aprofundamento e neste capítulo estamos somente versando sobre noções básicas necessárias para abordarmos as medidas de tendência central.

Até esse momento estamos discutindo a coleta de dados, representando apenas o começo no que diz respeito à análise estatística. Partimos assim, para a

organização dos dados e o primeiro passo é a construção da **DISTRIBUIÇÃO DE FREQUÊNCIAS** em forma de tabela.

Suponha que um pesquisador esteja interessado em identificar qual é o gosto pela Matemática e pela Língua Portuguesa de 200 alunos matriculados na 7ª série da “Escola Ratatui”. Após a coleta de dados o pesquisador terá em mão um banco de dados brutos onde encontramos a opção de cada aluno sobre seu gosto por Matemática e Língua Portuguesa.

Tabela B.26: Gosto pela Matemática e por Língua Portuguesa de 200 alunos matriculados na 7ª série da Escola Ratatui

Identificação	Gosto pela Matemática				Gosto por Língua Portuguesa			
	Não	Pouco	Regular	Muito	Não	Pouco	Regular	Muito
Sujeito 1	X							X
Sujeito 2		X					X	
Sujeito 3	X				X			
Sujeito 4			X				X	
Sujeito 5	X							X
Sujeito 6			X		X			
Sujeito 7		X					X	
Sujeito 8				X				X
Sujeito 9			X				X	
Sujeito 10			X		X			
(...)								
Sujeito 200	X					X		

Quando fazemos a opção de estudar uma variável de cada vez, ou seja, estudar a variável gosto pela Matemática e depois a variável gosto por Língua Portuguesa, iremos obter uma **análise unidimensional**. Um método para facilitar a análise unidimensional desses dados brutos é observar e anotar a quantidade de vezes que aquelas variáveis apareceram e construir uma distribuição de frequências. Para este estudo, adotaremos a definição de **DISTRIBUIÇÃO DE FREQUÊNCIAS** como sendo “*essa relação estabelecida, na qual cada opção tem apenas um valor de observações feitas e todas as opções são associadas a um único valor*” (NOVAES; COUTINHO, 2008, p. 22).

Para cada variável e dependendo do objetivo da análise dos dados a distribuição de frequências pode ser representada por gráficos ou tabelas.

Tabela B.27: Distribuição de 200 alunos quanto ao gosto pela Matemática, adaptado de CAZORLA e SANTANA, 2006, p. 15.

Gosto pela Matemática	Nº de Pessoas	
	( <i>f</i> )	(%)
Não	32	16,0
Pouco	48	24,0
Regular	96	48,0
Muito	24	12,0
Total	200	100,0

Tabela B.28: Distribuição de 200 alunos quanto ao gosto pela Língua Portuguesa, adaptado de CAZORLA e SANTANA, 2006, p. 15.

Gosto pela Língua Portuguesa	Nº de Pessoas	
	( <i>f</i> )	(%)
Não	24	12,0
Pouco	48	24,0
Regular	90	45,0
Muito	38	19,0
Total	200	100,0

Nessas tabelas de distribuição de freqüências de dados nominais, podemos notar três colunas: a da esquerda indica as opções possíveis para a variável “gosto pela Matemática” ou “gosto pela Língua Portuguesa”; a coluna adjacente indica o número de vezes que cada opção apareceu e recebe o nome de **freqüência absoluta** ou (***f***); finalmente a coluna (**%**) ou **freqüência relativa** representa o valor de *f* expresso em porcentagem. Na última linha da tabela encontramos o total (***N***) de alunos pesquisados, ou seja, 200 alunos. É importante ressaltar que utilizamos a notação ***N*** quando nos referimos ao tamanho da população e a notação ***n*** quando estamos trabalhando com uma amostra.

Uma rápida observação nas distribuições de freqüências das tabelas 2 e 3 revelam claramente que um grande número de alunos dentre os 200 pesquisados, se sentem indiferentes ao gosto tanto pela Matemática quanto à Língua Portuguesa.

Para a representação gráfica é necessário que consideremos a natureza dos dados e o objetivo da análise. Quando trabalhamos com **variáveis qualitativas ou quantitativas de dados discretos**, podemos lançar mão de gráficos de setor (quando o objetivo é ter uma visão do tipo “parte/todo”) ou gráficos de barras horizontais ou colunas, também chamado de barras verticais (quando o objetivo é comparação entre as partes).

No gráfico 1 encontramos a representação da variável qualitativa discreta “gosto pela Matemática” como gráfico de setor pelo motivo de favorecer o entendimento do tipo “parte-todo” e por ter um maior apelo visual. Para Levin e Fox (2007) as tabelas e quadros podem evocar temor, tédio, apatia ou mesmo desentendimento, em contrapartida as representações gráficas ou ilustrativas despertam nas pessoas mais atenção.

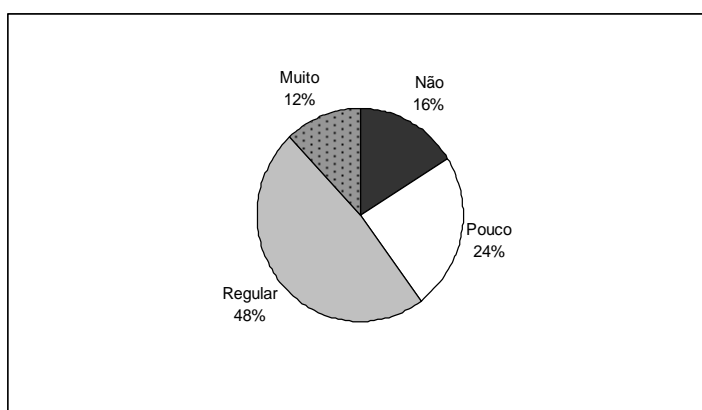


Figura B.71: Gráfico de Setores da distribuição de 200 alunos quanto ao gosto por Matemática, retirado de CAZORLA e SANTANA, 2006, p. 15.

Observando o gráfico 1, notamos que o número de alunos que tem “regular” gosto pela Matemática, nos salta aos olhos, favorecendo a comparação parte/todo: quase a metade do todo pesquisado (48%) optou por regular; os que gostam pouco representam quase  $\frac{1}{4}$  do total pesquisado, (24%) ; os que gostam muito representam a “fatia” menor do todo (12%), enquanto os que não gostam de Matemática (16%) representam  $\frac{1}{3}$  dos que optaram por regular (48%). Analisando os que optaram por “muito” (12%) com os que optaram por “não” (16%) podemos perceber que juntos eles perfazem um pouco mais que  $\frac{1}{4}$  dos entrevistados.

Nesse momento, não pretendemos inferir sobre esses gráficos e tabelas, a intenção é somente ilustrar situações hipotéticas para entendermos quando usar um ou outro tipo de gráfico.

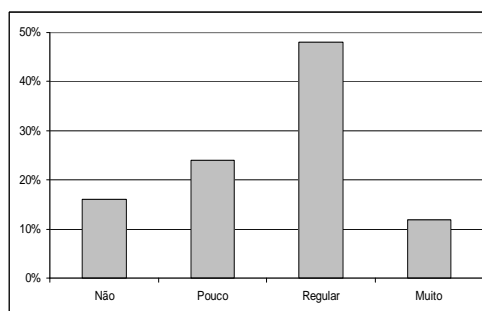


Figura 72: Gráfico de Colunas da distribuição de 200 alunos quanto ao gosto por Matemática, retirado de CAZORLA e SANTANA, 2006, p. 17.

Como já dito antes, esse gráfico pode ser construído com barras horizontais, fato esse que, em alguns casos, pode alterar a visualização da informação. A construção de um gráfico de barras verticais ou colunas são feitos de acordo com uma disposição padrão. Por exemplo, no gráfico 2, no eixo vertical (das ordenadas) anotamos a frequência relativa, pois facilita a observação de padrões de comportamento; no eixo horizontal (abscissas) anotamos as categorias.

Os gráficos de barras verticais ou colunas também oferecem a alternativa de combinar mais de uma variável, como, por exemplo, a distribuição do gosto pela Matemática em comparação com o gosto por Língua Portuguesa dos pesquisados.

Tabela B.29: Relação entre gosto pela Matemática e gosto por Língua Portuguesa, baseado em CAZORLA e SANTANA, 2006, p. 17.

Gosto por	Nº de pessoas			
	Matemática		Língua Portuguesa	
	(f)	(%)	(f)	(%)
Não	32	16,0	24	12,0
Pouco	48	24,0	48	24,0
Regular	96	48,0	90	45,0
Muito	24	12,0	38	19,0
Total	200	100,0	200	100,0

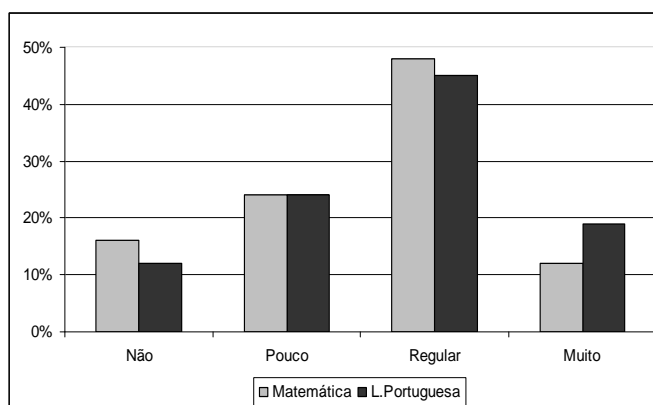


Figura B.73: Gosto por Matemática e Língua Portuguesa, baseado em CAZORLA e SANTANA, 2006, p. 18.

Analisando a tabela 4 e o gráfico 3, observamos que existe o mesmo número de alunos que dizem gostar “pouco” de Matemática e de Língua Portuguesa. Temos quase que metade dos pesquisados respondendo “regular” para as duas disciplinas, porém Matemática recebeu mais votos que Língua Portuguesa. Com relação a “não” gostar de uma das disciplinas, Matemática é menos apreciada uma vez que obteve mais votos que Língua Portuguesa e com relação a gostar “muito”, Matemática continua sendo a menos apreciada, uma vez que Língua Portuguesa obteve mais votos.

Agora, quando trabalhamos tabelas com **variáveis quantitativas contínuas**, utilizamos a distribuição de dados agrupados em classes, fato esse, que favorece a identificação de intervalos onde os valores observados estão inseridos.

Para que haja agrupamentos definimos o número de classes (**K**) bem como a amplitude<sup>7</sup> (**h**) das classes. O número de classes de **agrupamentos pré-determinados**, via de regra, não é nem inferior a 5 nem superior a 10 e devemos sempre observar se essa escolha não irá interferir na clareza da análise, muito menos no aparecimento de classe com freqüência zero.

Tomemos como exemplo, a tabela 5 que apresenta a organização da idade de cento e cinqüenta e três pessoas que compraram livros na Livraria “Euleio” durante o mês de janeiro de 2008:

Tabela B.30: Distribuição de freqüências das idades de 153 pessoas que compraram livros na Livraria Euleio no mês de jan de 2008, baseado em NOVAES e COUTINHO, 2008, p. 22.

Idade	(f)	Idade	(f)	Idade	(f)	Idade	(f)	Idade	(f)
6	0	21	3	36	2	51	2	66	0
7	1	22	5	37	11	52	1	67	0
8	0	23	0	38	0	53	1	68	1
9	2	24	2	39	1	54	0	69	0
10	1	25	15	40	5	55	4	70	0
11	0	26	13	41	4	56	8	71	1
12	0	27	9	42	0	57	0	72	0
13	4	28	3	43	0	58	0	73	0
14	0	29	8	44	2	59	0	74	0
15	0	30	7	45	8	60	0	75	0
16	1	31	2	46	3	61	1	76	1
17	0	32	0	47	0	62	0	77	0
18	1	33	4	48	0	63	0	78	0
19	0	34	1	49	1	64	0	79	0
20	4	35	6	50	2	65	2	80	0
TOTAL	14		78		39		19		3

Como a maior idade observada é 76 anos e a menor 7 anos, podemos utilizar a amplitude 10 ( $h=10$ ) que ainda sim ficaremos com 7 classes e estaremos dentro do limite convencional. Podemos começar a contagem das classes pelo menor valor observado, mas também podemos utilizar um valor menor ainda criando assim, classes que englobem todas as idades. Em nosso exemplo, a menor idade é 7 anos, porém, iremos começar nosso intervalo por 6 anos, tomando valores de 6 a 76, incluindo assim os valores observados de 7 a 76 anos. “Essa opção permite, em alguns casos, criar classes com extremos que propiciem uma leitura mais fácil. Tal escolha depende dos objetivos e do grau de precisão exigidos” (NOVAES; COUTINHO, 2008, p. 26)..

<sup>7</sup> A amplitude de uma classe, ou amplitude intervalar, é uma medida dada pela diferença entre o maior e o menor valor dessa classe. (NOVAES; COUTINHO, 2008).

Tabela B.31: Distribuição de 153 pessoas segundo a faixa etária, baseado em NOVAES e COUTINHO, 2008, p. 26.

Idade (anos)	Nº de Pessoas	
	( $f_i$ )	(%)
06   16	8	5,23
16   26	31	20,26
26   36	53	34,64
36   46	33	21,57
46   56	14	9,15
56   66	11	7,19
66   76	3	1,96
Total	153	100,00

Para esse tipo de distribuição, de dados agrupados em classes, podemos construir um gráfico que recebe o nome de **histograma**.

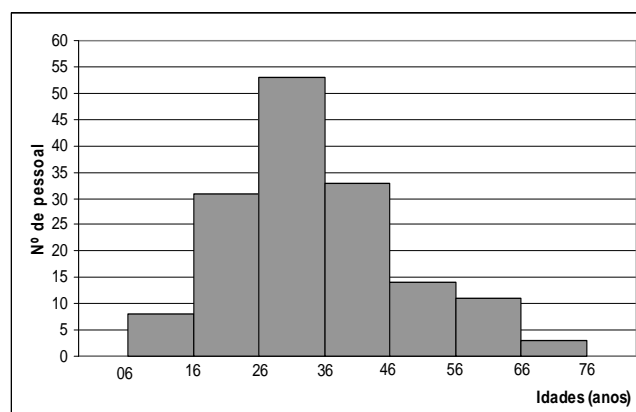


Figura 74: Histograma, baseado em NOVAES e COUTINHO, 2008, p. 26.

Se localizarmos geometricamente pontos que definem uma curva - pontos esses que têm como coordenadas o ponto médio de cada classe (abscissa) e a frequência observada (ordenada) – e uni-los por segmentos de reta obteremos o **polígono de freqüências**. No gráfico 5 observamos a localização dos pontos que possibilitarão a construção dos segmentos de reta:

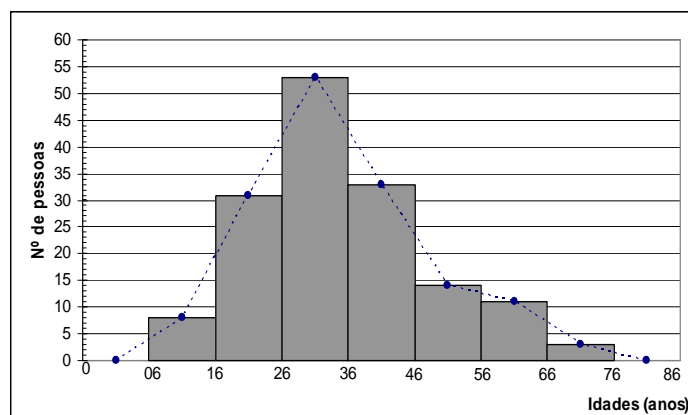


Figura B.75: Polígono de Freqüências e Histograma, baseado em NOVAES e COUTINHO, 2008, p. 27.

O polígono de freqüências não intenciona evidenciar as diferenças entre as variáveis, mas sim tende a dar ênfase na evolução dessa e sua continuidade ao longo de uma escala. As duas representações gráficas, histograma e polígono de freqüências, se equivalem e para Novaes e Coutinho (2008) a “*equivalência entre o histograma e o polígono de freqüências é garantida pela igualdade entre as medidas das áreas abaixo da linha que define o polígono de freqüências e no interior do histograma*”

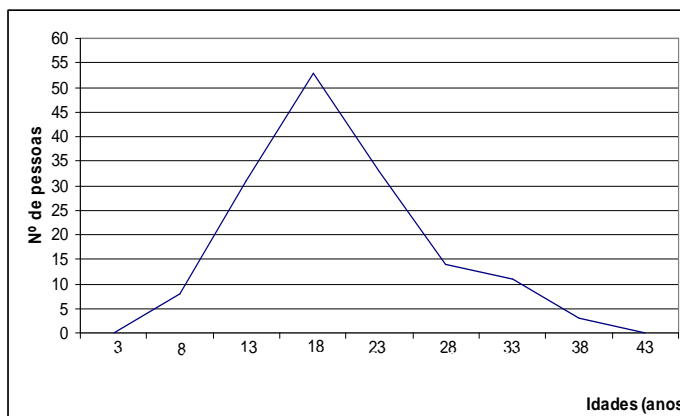


Figura B.76: Polígono de Freqüências, baseado em NOVAES e COUTINHO, 2008, p. 27.

Até esse momento, discutimos o agrupamento das variáveis quantitativas contínuas quando são pré-definidas suas classes e amplitude, agora, caso não tenhamos nenhum valor já estabelecido para assumir esses valores e imaginando que se deseje criar classes com a mesma amplitude, temos a **Regra de Sturges** ou o **Crítério da raiz** que calcula um número aproximado de quantidade de classes e de amplitude:

- Cálculo do número de classes  $K$

<b>Regra de Sturges</b>	<b>Crítério da raiz</b>
$K \cong 1 + 3,22 \cdot \log n$	$n \leq 25$ , usamos $K = 5$
onde	Ou
$K = \text{número de classes}$	$n \geq 25$ , usamos $k \cong \sqrt{n}$
$n = \text{número de elementos da amostra}$	

Quadro B.6: Fórmula da Regra de Sturges e do Crítério da raiz para cálculo do número de classes em distribuição de freqüências agrupadas, retirado de NOVAES e COUTINHO, 2008, p. 27.

- Cálculo da amplitude (h) dos intervalos

$$h = \frac{x_{\text{máximo}} - x_{\text{mínimo}}}{K}, \quad \text{onde } A_t = x_{\text{máximo}} - x_{\text{mínimo}}, \quad \text{então } h = \frac{A_t}{K}$$

Quadro B.7: Fórmula para calcular a amplitude dos intervalos em distribuição de frequências agrupadas, retirado de NOVAES e COUTINHO, 2008, p. 27.

Vamos retornar ao exemplo das cento e cinquenta e três pessoas que compraram livros na Livraria “Euleio” durante o mês de janeiro de 2008. Para que seja possível o cálculo do número de classes com a mesma amplitude, iremos utilizar a Regra de Sturges e o cálculo da amplitude intervalar também chamado de amplitude das classes:

- Cálculo do Número de classes - 
$$\begin{cases} k \cong 1 + 3,22 \cdot \log 153 \\ k \cong 1 + 3,22 \cdot 2,1847 \quad (\log 153 \cong 2,18469) \\ k \cong 8 \end{cases}$$
- Cálculo da Amplitude total ( $A_t$ ) - 
$$\begin{cases} A_t = 77 - 7 \\ A_t = 70 \end{cases}$$
- Cálculo da Amplitude Intervalar - 
$$\begin{cases} h = \frac{A_t}{k} \\ h = \frac{70}{8} \quad h = 8,75 \\ h = 8,75 \end{cases}$$

Na prática, tomamos o valor da amplitude intervalar sempre com a mesma formatação dos dados coletados. Como estamos trabalhando com um exemplo de distribuição de idades expressas em valores exatos, podemos arredondar o valor encontrado  $h = 8,75$  para  **$h = 9$** .

Então, com  **$K = 8$**  e  **$h = 9$** , podemos construir outra tabela de distribuição de frequências para a variável “idade”:

Tabela B.32: Distribuição de 153 pessoas segundo a idade, expressa em anos, baseado em NOVAES e COUTINHO, 2008, p. 28.

Idade (anos)	Nº de pessoas	
	(f)	(%)
07 - 16	8	5,2
16 - 25	16	10,5
25 - 34	61	39,9
34 - 43	30	19,6
43 - 52	18	11,8
52 - 61	14	9,1
61 -70	4	2,6
70 - 79	2	1,3
Total	153	100,0

As chamadas variáveis quantitativas contínuas também podem ser representadas pelo **Diagrama de Ramo e Folha**. Vamos considerar as notas que os 25 alunos da 7ª A da Escola Ratatui tiveram no 2º bimestre do ano de 2001, para tanto observe o quadro 3 que contém essas notas.

Nome do aluno	Nota 2º bimestre	Nome do aluno	Nota 2º bimestre
Amanda	2,0	Daniel	6,5
Amauri	4,0	Danilo	7,0
Aurora	5,0	Fernando	6,5
Armando	6,5	Francisco	7,5
Bia	7,0	Fernanda	7,5
Benedito	4,0	Guilherme	8,5
Bianca	6,0	Heitor	7,5
Bruno	5,5	José	7,0
Caíque	4,5	Mauricio	7,5
César	8,5	Milena	9,5
Camila	9,0	Pedro	9,0
Cintia	6,5	Paula	10,0
Damares	4,5		

Quadro B.8: Quadro das notas do 2º bimestre do ano de 2001 de 25 alunos da 7ª A da escola Ratatui, adaptado de CAZORLA e SANTANA, 2006, p.25.

Com base nesse quadro, construímos a tabela de distribuição de freqüências com vistas na comparação entre a tabela e o diagrama de ramo e folha.

Tabela B.33: Distribuição de freqüência das notas no 2º bimestre do ano de 2001 de 25 alunos da 7ª A da escola Ratatui, baseado em CAZORLA e SANTANA, 2006, p.25.

Notas	Nº de alunos	
	(f)	(%)
2,0 - 3,0	1	4,0
3,0 - 4,0	0	0,0
4,0 - 5,0	4	16,0
5,0 - 6,0	2	8,0
6,0 - 7,0	5	20,0
7,0 - 8,0	7	28,0
8,0 -9,0	2	8,0
9,0 - 10,0	4	16,0
Total	25	100,0

Ramo	Folha
2	0
3	
4	0055
5	05
6	05555
7	0005555
8	55
9	005
10	0

Figura B.77: Diagrama de ramo e folha das notas do 2º bimestre do ano de 2001 de 25 alunos da 7ª A da escola Ratatui, baseado em CAZORLA e SANTANA, 2006, p.25.

No diagrama de ramo e folha, o ramo representa as unidades inteiras e as folhas, os decimais. Quando observamos que uma “folha” está vazia, significa que não observamos a ocorrência daquela variável e não será preenchida com zero por que isso indicaria a freqüência (1) daquela variável. Nesse tipo de representação temos todas as informações necessárias, uma vez que conseguimos observar exatamente a freqüência que a variável apareceu e o valor específico de cada uma dentro da classe, já na representação tabular, sabemos apenas a quantidade de variáveis que aparecem em uma classe.

A representação tabular é importante para o entendimento e apresentação dos dados, porém, muitos pesquisadores e autores de informações estatísticas recorrem a auxílios visuais – como gráficos em setores, barras, colunas, linhas e polígonos de freqüências - pelo motivo de ser uma rápida visualização de um conjunto de dados proporcionando melhor legibilidade de seus resultados. Além disso, podemos lançar mão de outras formas de representação do conjunto de dados: as medidas de tendência central.

A estimativa do cálculo da moda para a distribuição de freqüências com dados agrupados é feita por meio de semelhança de triângulos ou a **fórmula de Czuber**. Para tal estimativa devemos localizar a classe com maior número de elementos (classe modal), a classe anterior e a posterior a ela ou lançar mão da representação da distribuição no histograma. Cf. NOVAES, D.V.; COUTINHO, C.Q.S., Estatística para Educação Profissional, p. 46.

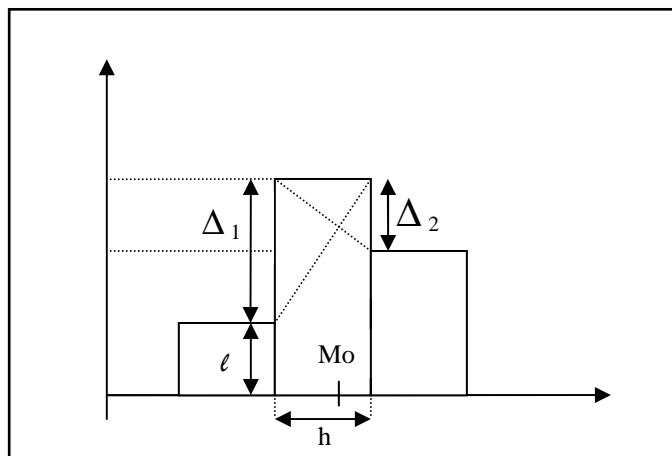


Figura C.78: Gráfico demonstrativo do cálculo da Moda em dados agrupados em intervalos, retirado de NOVAES e COUTINHO, 2008, p.47.

Semelhança de triângulos	Fórmula de Czuber
$\frac{\Delta_1}{\Delta_2} = \frac{x}{h-x}$ $\Delta_2 x = \Delta_1 (h-x)$ $\Delta_1 x + \Delta_2 x = \Delta_1 h$ $x(\Delta_1 + \Delta_2) = \Delta_1 h$ $x = \frac{\Delta_1 h}{\Delta_1 + \Delta_2}$ $mo = l + x$	$mo = l + \frac{\Delta_1 h}{\Delta_1 + \Delta_2}$

Quadro C.9: Demonstração do cálculo da Moda por meio da Semelhança de Triângulos e Fórmula de Czuber, retirado de NOVAES e COUTINHO, 2008, p. 47.

Onde:

$\Delta_1$  – diferença entre a frequência da classe modal e a frequência da classe anterior;

$\Delta_2$  – diferença entre a frequência da classe modal e frequência da classe posterior;

h – amplitude da classe modal;

l – limite inferior da classe modal.

Considere que a fábrica de autopeças “Façobem” estimou o tempo de vida útil da peça XY e obteve a distribuição da tabela 10:

Tabela C.34: Distribuição do tempo de vida útil da peça XY produzida pela fábrica de autopeças “Façobem”, adaptado de NOVAES e COUTINHO, 2008, p. 44.

Nº de meses (vida útil)	Nº de peças (f)
0 - 12	06
12 - 24	43
24 - 36	91
36 - 48	132
48 - 60	84
60 - 72	55
72 -84	09
Total	420

Primeiramente devemos localizar a classe modal, 36|- 48 com frequência 132, a frequência da classe anterior ( $f = 91$ ) e da posterior ( $f = 84$ ), em seguida utilizando a representação geométrica da moda no histograma iremos estimar o valor da moda:

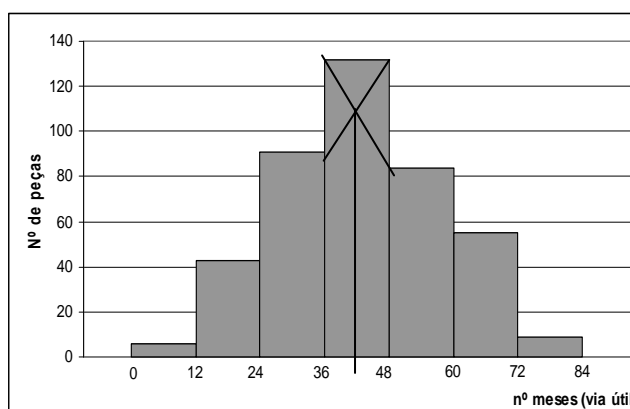


Figura C.79: Histograma da distribuição do tempo de vida útil da peça XY produzida pela fábrica de autopeças “Façobem”, baseado em NOVAES e COUTINHO, 2008, p. 46.

Partiremos assim para a estimativa da moda:

Semelhança de Triângulos	Fórmula de Czuber
$\frac{132-91}{132-84} = \frac{x}{12-x}$	
$\frac{41}{48} = \frac{x}{12-x}$	$mo = 36 + \frac{(132-91)12}{41+48}$
$48x = 41(12-x)$	$mo = 36 + 5,53$
$48x + 41x = 492$	$mo = 41,53$
$x(48+41) = 492$	
$x = \frac{492}{89}$	
$x = 5,53$	
$mo = 36 + 5,53$	
$mo = 41,53$	

Quadro C.10: Cálculo da moda da distribuição de dados intervalares do tempo de vida útil da peça XY produzida pela fábrica de autopeças “Façobem”, baseado em NOVAES e COUTINHO, 2008, p. 46.

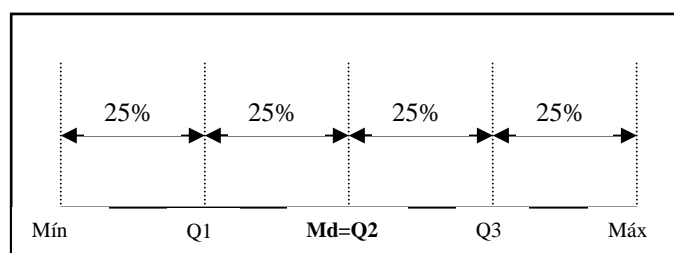
Para a distribuição do tempo de vida útil da peça XY produzida pela fábrica de autopeças “Façobem”, encontramos  $mo = 41,53$ .

## APÊNCIDE D

### A MEDIANA

Como já dissemos no capítulo II, a mediana também pode ser classificada como uma das medidas separatrizes<sup>8</sup>. Para esta pesquisa trataremos apenas dos quartis - divisão de um conjunto de dados em quatro partes iguais:

- primeiro quartil ( $Q_1$ ) – é o número que limita os valores do conjunto que são  $\leq 25\%$  de  $N$ ;
- segundo quartil ( $Q_2=Md$ ) – é o número que limita os valores do conjunto que são  $\leq 50\%$  de  $N$ ;
- terceiro quartil ( $Q_3$ ) – é o número que limita os valores do conjunto que são  $\leq 75\%$  de  $N$



Quadro D.11: Demonstração de determinação dos quartis, retirado de NOVAES e COUTINHO, 2008, p. 32.

Para encontrar a mediana ( $Md$ ) ou o segundo quartil ( $Q_2$ ) de **variáveis quantitativas discretas** representadas em tabelas, devemos construir uma coluna auxiliar onde registramos a freqüência acumulada crescente ( $F_{ac}$ ), o que facilitará a localização da posição de  $Q_2$ .

Tomemos como exemplo o cálculo da mediana da tabela que apresenta a distribuição de um determinado equipamento segundo o tempo (em meses) após o qual estes precisaram de reparos.

<sup>8</sup> As medidas separatrizes representam uma divisão dos dados, de acordo com o interesse da pesquisa, em quantas partes desejamos. Encontramos as divisões em: quartis (divisão em quatro partes iguais), os decis (dez partes iguais) e os percentis (cem partes iguais).

Tabela D.35: N° de equipamentos distribuídos segundo o tempo (em meses) após o qual estes precisaram de reparos. Retirado de NOVAES e COUTINHO, 2008, p. 33.

Duração antes de necessidade de reparo (em meses)	N° equipamentos ( $f_i$ )	Frequência acumula crescente ( $F_{ac}$ )
06	01	01
09	05	06
12	14	20
15	09	29
18	15	44
24	16	60
Total	60	

Como temos 60 equipamentos, devemos dividir esse número em 4 partes iguais, ou seja,  $\frac{60}{4} = 15$  (**divisão exata**); o número encontrado significa quantos elementos terá cada uma das partes.

15 elementos 1°	15 elementos 15°	15 elementos 16°	15 elementos 30°	15 elementos 31°	15 elementos 45°	15 elementos 46°	15 elementos 60°
Q <sub>1</sub>		Q <sub>2</sub> = Md		Q <sub>3</sub>			

Quadro D.12: Demonstração de determinação dos quartis retirado de NOVAES e COUTINHO, 2008, p. 33

Assim, podemos constatar que a mediana para essa distribuição está entre a 30ª e 31ª posição e para encontrar esse valor, basta buscar na coluna ( $F_{ac}$ ) quais são os elementos que estão nessas posições e calcular a média aritmética entre eles.

$$Q_2 (Md) = \frac{18+18}{2} = 18 \text{ meses}$$

Analogamente, encontramos os valores do primeiro ( $Q_1$ ) e terceiro quartil ( $Q_3$ ). Porém quando o número de elementos pesquisados deixa **resto 1** na divisão por quatro, a mediana pode ser determinada da seguinte forma:

Tabela D.36: N° de equipamentos distribuídos segundo o tempo (em meses) após o qual estes precisaram de reparos. Retirado de NOVAES e COUTINHO, 2008, p. 34.

Duração antes de necessidade de reparo (em meses)	N° equipamentos ( $f_i$ )	Frequência acumula crescente ( $F_{ac}$ )
06	01	01
09	05	06
12	14	20
15	10	30
18	15	45
24	16	61
Total	61	

Nesse caso, como temos 61 equipamentos e devemos dividi-los em quatro partes iguais, não teremos uma divisão exata, mas sim, resto igual a 1:  $\frac{61}{4} = 15,25$ . Para tanto, continuaremos distribuindo 15 elementos em cada parte, porém esta sobra de um elemento será a posição da mediana.

15 elementos 1°	15 elementos 15°	1 31°	15 elementos 32°	15 elementos 46°	15 elementos 47°	61°
Q <sub>1</sub>		Q <sub>2</sub> = Md	Q <sub>3</sub>			

Quadro D.13: Demonstração de determinação dos quartis baseado em NOVAES e COUTINHO, 2008, p. 34

Assim, o primeiro quartil (Q<sub>1</sub>) será a média aritmética entre o 15° e 16° elemento; a mediana é o 31° elemento da distribuição, ou seja, **Md = 18 meses** e o terceiro quartil (Q<sub>3</sub>) será a média aritmética entre 46° e 47° elemento.

Para tabelas cujo número de elementos pesquisados, ao ser dividido por quatro deixa **resto 2**, procedemos da seguinte forma:

Tabela D.37: N° de equipamentos distribuídos segundo o tempo (em meses) após o qual estes precisaram de reparos. Retirado de NOVAES e COUTINHO, 2008, p. 35.

Duração antes de necessidade de reparo (em meses)	N° equipamentos ( $f_i$ )	Frequência acumula crescente ( $F_{ac}$ )
06	01	01
09	05	06
12	14	20
15	11	31
18	15	46
24	16	62
Total	62	

Para essa distribuição temos 62 elementos e este número dividido em quatro partes deixa resto 2:  $\frac{62}{4} = 15,5$ . Continuamos distribuindo 15 elementos em cada parte e o resto 2, será inserido da seguinte forma: um elemento entre a primeira e segunda parte e o outro entre a terceira e quarta parte.

15 elementos 1°      15°	1 16°	15 elementos 17°      31°	15 elementos 32°      46°	1 47°	15 elementos 48°      62°
Q <sub>1</sub>		Q <sub>2</sub> = Md		Q <sub>3</sub>	

Quadro D.14: Demonstração de determinação dos quartis baseado em NOVAES e COUTINHO, 2008, p. 35

Desse jeito o primeiro quartil é o 16° elemento da distribuição; procurando na coluna ( $F_{ac}$ ) constatamos que  $Q_1 = 12$  meses; a mediana é o valor da média aritmética entre o 31° e o 32° elemento ( $Md = Q_2 = \frac{15+18}{2} = 16,5$  meses) e o terceiro quartil é o 47° elemento,  $Q_3 = 24$  meses.

Já para distribuições com um total de elementos que dividido em quatro partes resulta em **resto 3**, observe como iremos proceder para encontrar a mediana:

Tabela D.38: N° de equipamentos distribuídos segundo o tempo (em meses) após o qual estes precisaram de reparos. Retirado de NOVAES e COUTINHO, 2008, p. 35

Duração antes de necessidade de reparo (em meses)	N° equipamentos ( $f_i$ )	Frequência acumula crescente ( $F_{ac}$ )
06	01	01
09	05	06
12	14	20
15	12	32
18	15	47
24	16	63
Total	63	

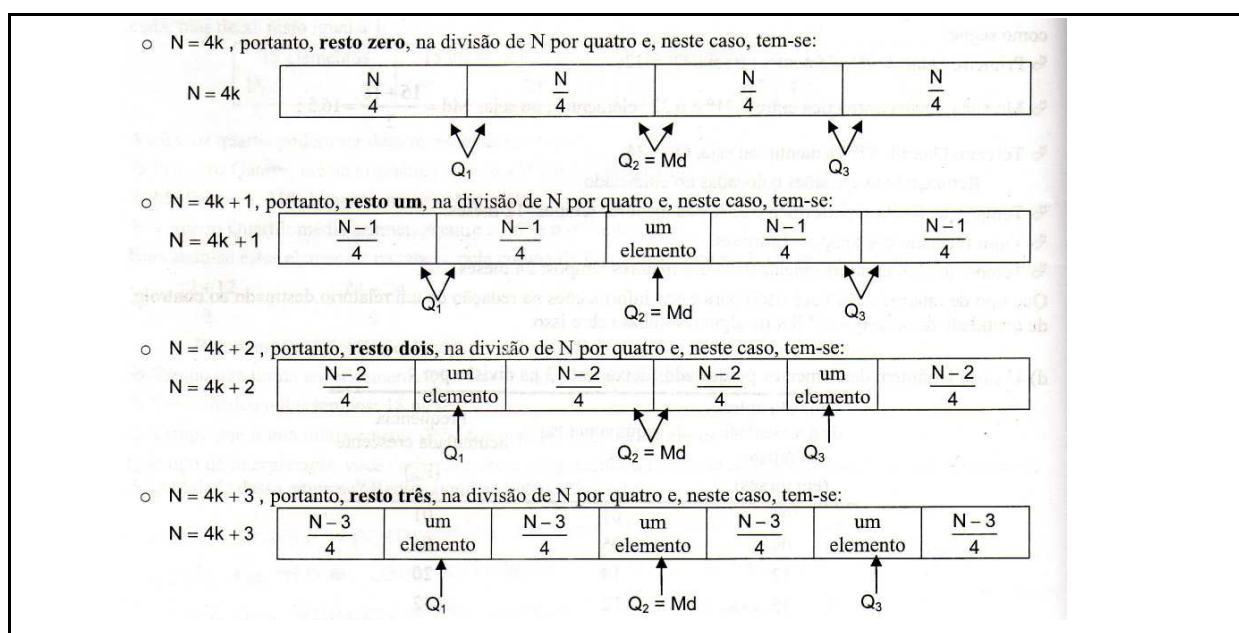
Para esse caso temos a seguinte distribuição de elementos em cada parte:  
 $\frac{63}{4} = 15,75$ , ou seja, cada quartil abriga 15 elementos e os 3 elementos que sobram, devem ser inseridos entre os quartis, a saber:

15 elementos 1° 15°	1 16°	15 elementos 17° 31°	1 32°	15 elementos 33° 47°	1 48°	15 elementos 49° 63°
Q <sub>1</sub>		Md=Q <sub>2</sub>		Q <sub>3</sub>		

Quadro D.15: Demonstração de determinação dos quartis baseado em NOVAES e COUTINHO, 2008, p. 36

Assim, os quartis serão determinados da seguinte forma: o primeiro quartil é o 16º elemento ( $Q_1=12$  meses), a mediana será o 32º elemento ( $Q_2=Md=15$  meses) e o terceiro quartil será o 48º elemento ( $Q_3=24$  meses).

Apresentaremos a seguir o esquema de determinação dos quartis e mediana:



Quadro D.16: Esquema da determinação dos quartis, retirado de NOVAES e COUTINHO, 2008, p. 36.

Considere agora uma distribuição de variável quantitativa contínua onde podemos encontrar uma tabela que representa as idades de consumidores que usam certo tipo de celular:

Tabela D.39: Distribuição das pessoas que usam um determinado tipo de celular segundo as idades (anos), retirado de NOVAES e COUTINHO, 2008, p. 37.

Idade (anos)	Nº de pessoas		Colunas auxiliares
	$(f_i)$	$(Fr_{ac})$	Posições
14   - 22	5	0,0806	1° ao 5°
22   - 30	5	0,1612	6° ao 10°
30   - 38	4	0,2257	11° ao 14°
<b>38   - 46</b>	<b>14</b>	<b>0,4515</b>	<b>15° ao 28°</b>
<b>46   - 54</b>	<b>19</b>	<b>0,758</b>	<b>29° ao 47°</b>
54   - 62	7	0,8709	48° ao 54°
62   - 70	8	0,9999	55° ao 62°
Total	62		

Quando temos os dados agrupados em intervalos, estimamos o valor procurado e interpolamos o mesmo, para tanto, consultamos a coluna auxiliar de freqüências relativas acumuladas e procuramos qual é a classe em que se encontra o  $Q_1$ ,  $Q_2=Md$  e  $Q_3$ .

Ao observarmos a tabela 15, notamos que 0,25 dos valores estão localizados na classe de 38 – 46 anos. Agora, “*por meio de uma regra de três simples, fazemos uma interpolação que leva em conta as diferenças entre os termos correspondentes das três colunas ...*” NOVAES e COUTINHO, 2008, p. 37.

Limite inferior da classe	$Q_1$	Limite superior da classe
$Fr_{ac}$ até o limite inferior	0,25	$Fr_{ac}$ até o limite superior

Quadro D.17: Demonstração de determinação dos quartis adaptado de NOVAES e COUTINHO, 2008, p. 37.

$$\frac{Q_1 - 38}{0,25 - 0,2257} = \frac{46 - 38}{0,4515 - 0,2257}$$

$$(Q_1 - 38) \cdot 0,2258 = 0,0243 \cdot 8$$

$$0,2258Q_1 - 8,5804 = 0,1944$$

$$Q_1 = \frac{0,1944 + 8,5804}{0,2258}$$

$$Q_1 = 38,86$$

Obter como medida do primeiro quartil o valor 38,86 significa que 25% dos pesquisados têm entre 14 e aproximadamente 38,86 anos de idade.

Para determinar a mediana ( $Md$ ) ou  $Q_2$ , procedemos de forma análoga ao cálculo de  $Q_1$ , não esquecendo de localizar na coluna de freqüência relativa acumulada o intervalo que engloba 50% dos dados ou 0,5 dos dados. Observando a tabela 15, constatamos que este número se encontra no intervalo de 46 | 54 anos. Então:

Limite inferior da classe	$Md$	Limite superior da classe
$Fr_{ac}$ até o limite inferior	0,5	$Fr_{ac}$ até o limite superior

Quadro D.18: Demonstração de determinação dos quartis, retirado de NOVAES e COUTINHO, 2008, p. 37.

$$\frac{Md - 46}{0,5 - 0,4516} = \frac{54 - 46}{0,7851 - 0,4516}$$

$$(Md - 46) \cdot 0,3335 = 0,3872$$

$$0,3335Md - 15,341 = 0,3872$$

$$Md = \frac{0,3872 + 15,341}{0,3335}$$

$$Md \cong 47,1610$$

Esta medida, a mediana com valor aproximado de 47,16 representa a idade máxima que iremos encontrar a metade do número de pesquisados, em outras palavras, temos 50% dos entrevistados com idades entre 14 e 47,16 anos inclusive, e os outros 50% têm de aproximadamente 47,16 (inclusive) a 70 anos de idade. Cf. NOVAES, D.V.; COUTINHO, C.Q.S., Estatística para Educação Profissional, p. 38.

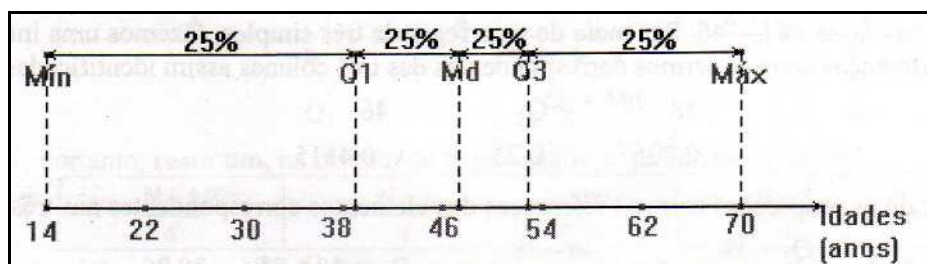
Finalmente, para o cálculo da idade que limita superiormente 75% dos entrevistados (Q3) devemos proceder da mesma forma, determinando as proporções entre as diferenças entre os elementos procurados:

$$\frac{Q3 - 46}{54 - 46} = \frac{0,75 - 0,4515}{0,758 - 0,4515} \Leftrightarrow Q3 = 46 + 7,7912 = 53,7$$

Quadro D.19: Demonstração de determinação dos quartis retirado de NOVAES e COUTINHO, 2008, p. 38.

O valor encontrado para Q3 foi de 53,7 anos, o que significa dizer que 75% dos entrevistados têm entre 14 e 53,7 anos (inclusive) e os 25% restantes têm de 53,7 até 70 anos.

Esquemáticamente temos:



Quadro D.20: Esquema da determinação dos quartis da distribuição das pessoas que usam um determinado tipo de celular segundo as idades (anos), retirado de NOVAES e COUTINHO, 2008, p. 38.

Estudando a densidade dos dados localizados em cada quartil, com vistas na variação dos dados, podemos perceber que 50% dos entrevistados estão concentrados na classe 38,3 |- 53,8 anos, em contrapartida os outros 50% estão mais dispersos nas classes de 14 |- 38,3 anos (25% dos entrevistados) e 53,8 |- 70 anos (25%). Esse tipo de estudo é de grande valia quando se quer estudar qual é a faixa de idade que provavelmente o lançamento de um produto será bem sucedido. Cf. NOVAES, D.V.; COUTINHO, C.Q.S., Estatística para Educação Profissional, p. 38.

Vamos considerar que um pesquisador realizou um teste com oito entrevistados e apurou o quociente de inteligência (QI) de cada um, representando em uma tabela:

Tabela E.40: QI de oito entrevistados, retirad de LEVIN e FOX, 2007, p. 82.

Entrevistados	QI
Sujeito 1	125
Sujeito 2	92
Sujeito 3	72
Sujeito 4	126
Sujeito 5	120
Sujeito 6	99
Sujeito 7	130
Sujeito 8	100
Total	864

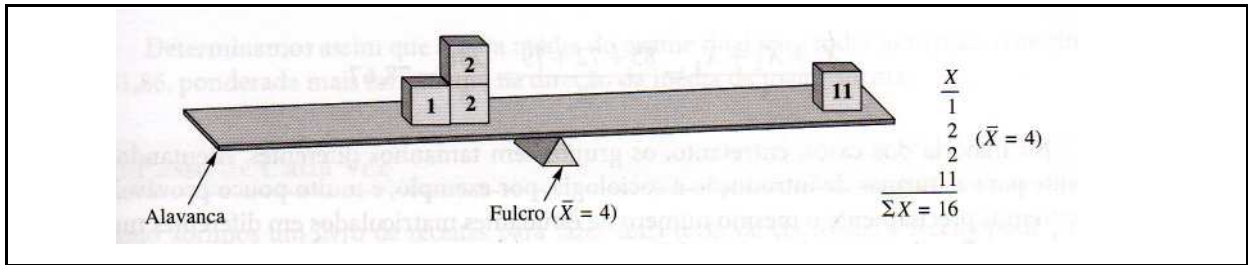
Aplicando a fórmula para calcular o QI médio dos oito entrevistados obtemos:

$$\bar{X} = \frac{125 + 92 + 72 + 126 + 120 + 99 + 130 + 100}{8} = \frac{864}{8} = 108$$

Analisando o valor da média aritmética de **distribuição simples** obtida, podemos perceber que este valor, ao contrário da moda, não é o que ocorre com mais freqüência e ao contrário da mediana ele não representa necessariamente o ponto médio de uma distribuição.

Fazendo uma comparação da média com uma gangorra ou alavanca, poderíamos dizer que este valor ( $\bar{X}$ ) faz o papel do fulcro e os valores das variáveis seriam os pesos de uma alavanca. Levin e Fox (2007) definem a média como sendo “o ponto de uma distribuição em torno do qual os valores acima dele se equilibram com os que estão abaixo”.

Para ilustrar a comparação da alavanca com a média e compreendermos o seu significado observe a figura do quadro 16:



Quadro E.21: Analogia da alavanca e fulcro para a média, retirado de LEVIN e FOX, 2007, p. 83.

Nessa ilustração, colocaram quatro pesos em uma alavanca de forma que ficassem equilibradas. Para que isso acontecesse foi necessário encontrar o exato local onde colocar o fulcro. Os pesos fazem o papel dos valores das variáveis e o fulcro o papel da média. O peso marcado 11 está 7 unidades à direita do fulcro enquanto o pesos marcados 1, 2 e 2 estão à esquerda do fulcro 3, 2 e 2 unidades respectivamente. Cf. LEVIN, J; FOX, J.A., Estatística para Ciências Humanas, p. 83.

Para que seja possível entender essa característica da média, primeiramente precisamos ter uma noção do papel do **desvio**. Levin e Fox (2007) definem como desvio o valor “*que indica a distância e a direção de qualquer escore bruto em relação à média*”. Para encontrar o desvio de uma variável, apenas subtraímos a média daquele valor:

$$\text{Desvio} = X - \bar{X} \quad \left\{ \begin{array}{l} \text{Onde} \\ X = \text{valor bruto arbitrário na distribuição} \\ \bar{X} = \text{média da distribuição} \end{array} \right.$$

Considere agora, o conjunto de dados brutos 9, 8, 6, 5 e 2 e o valor da média  $\left(\bar{X} = \frac{9+8+6+5+2}{5} = \frac{30}{5} = 6\right)$ . Utilizando o valor da média vamos calcular o desvio de cada variável:

Tabela E.41: Desvios de um conjunto de variáveis em torno da média  $\bar{X}$ , retirado de LEVIN e FOX, 2007, p. 83.

X	$X - \bar{X}$	
9	9-6=3	} +5
8	8-6=2	
6	6-6=0	
5	5-6=-1	} -5
2	2-6=-4	

Observe que a variável 6, também é o valor médio, analogamente ele representa o fulcro de uma balança, é o ponto de equilíbrio na distribuição. Se compararmos a soma dos desvios padrão (valor absoluto) acima da média com os abaixo da média veremos que são iguais. Esse fato ocorre quando a média da distribuição é o “centro de gravidade”.

Até esse momento discutimos a média aritmética de distribuição simples, mas, quando queremos encontrar a média de **dados agrupados** utilizamos a representação tabular com uma coluna auxiliar ( $f_i \cdot x_i$ ), onde  $x_i$  representa as variáveis estudadas,  $f_i$  significa a freqüência observada e  $n$  o total pesquisado.

Cálculo da Média Aritmética de dados agrupados	
Para medidas populacionais	Para medidas amostrais
$\mu = \frac{\sum_{i=1}^N f_i \cdot x_i}{N}$	$\bar{X} = \frac{\sum_{i=1}^n f_i \cdot x_i}{n}$

Quadro E.22: Demonstração das fórmulas para cálculos de média aritmética de dados agrupados, baseado em NOVAES e COUTINHO, 2008, p. 49.

Vamos considerar que queremos calcular o valor médio das diárias de pousadas da cidade de Vistajóia, escolhida pelos hóspedes no verão de 2000, conforme distribuição de freqüências da tabela 18:

Tabela E.42: Distribuição da diária de pousadas escolhidas pelos hóspedes da cidade de Vistajóia no verão de 2000, com coluna auxiliar para cálculo da média. Adaptado de NOVAES e COUTINHO, 2008, p. 49.

Diária (em reais)	Nº de pessoas	
	( $x_i$ )	Coluna auxiliar ( $f_i \cdot x_i$ )
50	45	45.50 = 2250
68	36	36.68 = 2448
95	13	13.95 = 1235
125	4	4.125 = 500
215	2	2.215 = 430
Total	100	6 863,00

Agora, aplicamos os valores encontrados na fórmula de  $\bar{X}$ ; note que estamos trabalhando com medidas amostrais:

$$\bar{X} = \frac{\sum_{i=1}^4 f_i \cdot x_i}{n}$$

$$\bar{X} = \frac{6863}{100} = 68,63$$

Assim, encontramos como valor médio das diárias R\$ 68,63, tal valor representa a medida em torno da qual as outras variam. Nesse momento não iremos abordar a questão da variação, item muito importante para a análise de dados, porém discutido logo mais.

Já para os dados agrupados em intervalos a forma de calcular é mesma, entretanto os valores de  $x_i$  correspondem aos pontos médios de cada uma das classes da distribuição em questão.

$$x_i = \frac{L_i + l_i}{2} \left\{ \begin{array}{l} \text{Onde} \\ L_i \text{ significa o limite superior da classe} \\ l_i \text{ significa o limite inferior da classe} \end{array} \right.$$

Para ilustrar esse cálculo, vamos voltar ao exemplo da distribuição das pessoas que usam um determinado tipo de celular segundo as idades em anos. Devemos construir uma tabela com colunas auxiliares, isso facilita os cálculos. Observe a tabela 20:

Tabela E.43: Distribuição das pessoas que usam um determinado tipo de celular segundo as idades (anos), retirado de NOVAES e COUTINHO, 2008, p. 37.

Idade (anos)	Nº de pessoas ( $f_i$ )	Colunas auxiliares	
		Ponto médio ( $x_i$ )	( $f_i \cdot x_i$ )
14   22	5	$(14 + 22) / 2 = 18$	90
22   30	5	26	130
30   38	4	34	136
38   46	14	42	588
46   54	19	50	950
54   62	7	58	406
62   70	8	66	528
Total	62	294	2828

Assim, partimos para o cálculo da média:  $\bar{X} = \frac{2828}{62} \cong 45,61$ . Esse número, 45,61 anos, representa a idade média dos 62 entrevistados. Para Novaes e Coutinho (2008) “o valor obtido no cálculo não deve ser arredondado, mas interpretado”. Ressaltamos que a média não indica a variação dos dados nem como esses dados estão concentrados em torno da média.

Temos também os dados representados graficamente e a forma de calcular a média não se modifica, apenas devemos prestar mais atenção nas informações contidas no gráfico e determinar os pontos médios de cada uma das classes associando-os a cada frequência correspondente.

A título de ilustração, vamos considerar a distribuição do tempo (minutos) que 16 entrevistados passaram ao telefone na espera do atendimento de uma empresa de telemarketing. Agora examine o gráfico dessa situação:

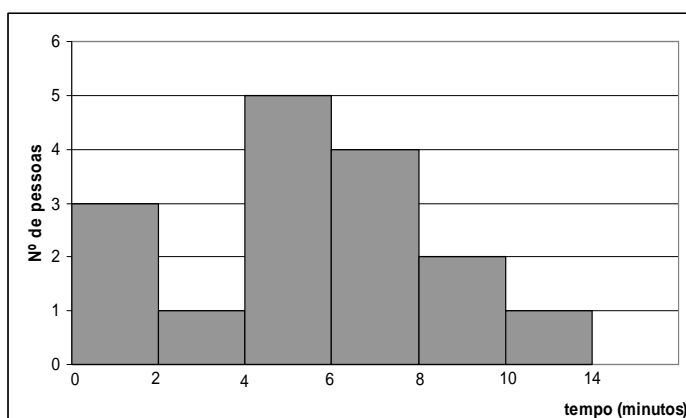


Figura E.80: Distribuição dos tempos (minutos) que 16 entrevistados passaram ao telefone na espera do atendimento de uma empresa de telemarketing, baseado em NOVAES e COUTINHO, 2008, p. 50.

O cálculo do ponto médio de classe aflora quase que intuitivamente e obtemos a frequência de cada classe somente observando o gráfico. Então, temos:

$$\bar{X} = \frac{1 \cdot 3 + 3 \cdot 1 + 5 \cdot 5 + 7 \cdot 4 + 9 \cdot 2 + 11 \cdot 1}{3 + 1 + 5 + 4 + 2 + 1} = \frac{88}{16} = 5,5 \text{ minutos}$$

Assim, as pessoas pesquisadas passam em média 5,5 minutos ao telefone sem ter atendimento pela empresa de telemarketing, número este que não indica a variação dos dados ao redor da média.

A variação dos dados é um assunto muito importante para dar conta da análise exploratória dos dados, por esse motivo, achamos necessário que se faça uma breve discussão sobre as medidas de variabilidade.

Até este momento estudamos as medidas que podem representar um conjunto de dados por meio de um único número médio ou típico, porém estes números empregados isoladamente nos dão uma visão parcial do todo, podendo até influenciar uma análise equivocada.

Encontramos o **desvio médio (dm)** que para Novaes e Coutinho (2008) “é a *media do valor absoluto dos desvios de cada valor em relação à média da distribuição*” e determinada pela fórmula:

$$dm = \frac{\sum_{i=1}^n |x_i - \bar{X}|}{n}$$

Esse processo de criação de uma medida de variabilidade, o desvio médio, apresenta a desvantagem de trabalhar com valores absolutos e que não admitem uma manipulação algébrica fácil. Como forma de tornar essa medida mais maleável, lançamos mão da **variância ( $s^2$ )** ou ( $\sigma^2$  **em caso de população**) que é calculada pela fórmula:

$$s^2 = \frac{\sum (x_i - \bar{X})^2}{n}$$

Para Levin e Fox (2007) “a *vantagem da variância sobre o desvio médio, além da natureza problemática dos valores absolutos, (...) é mais sensível ao grau de desvio na distribuição*”. Porém, alteramos a unidade de medida quando calculamos o seu quadrado e para que isso seja corrigido, tomamos a raiz quadrada da variância, o que nos dará o **desvio padrão (s)**. Então:

$$s = \sqrt{\frac{\sum (x_i - \bar{X})^2}{n}}$$

Sendo assim, Levin e Fox (2007) definem que o desvio padrão “*representa a variabilidade média em uma distribuição, porque mede a média dos desvios a contar*”

*da média*” e quanto maior for este número, maior é a variabilidade em torno da média de uma distribuição.