

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística

Exercícios Resolvidos em
Introdução à Estatística
para Ciências Sociais

Edna A. Reis

Relatório Técnico
RTE-03/2001

Relatório Técnico
Série Ensino

**Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Departamento de Estatística**

Exercícios Resolvidos em Introdução à Estatística para Ciências Sociais

Prof^a Edna Afonso Reis

com a colaboração da estudante
Viviane Alves Vieira

Primeira Edição – Setembro/2001

Prefácio

Esta apostila é o resultado de muitos semestres de experiência no ensino de Estatística Básica para alunos do curso de graduação em Ciências Sociais da UFMG. A apostila traz 51 exercícios divididos em dez seções e suas respectivas soluções comentadas. É indicada para a prática dos conceitos aprendidos nas aulas teóricas e discussão de tópicos específicos. Seu público alvo são os estudantes de graduação dos cursos da área de Ciências Humanas, especialmente para as Ciências Sociais.

Índice Geral

Primeira Parte: Enunciado dos Exercícios **9**

Seção 1: Abusos da Estatística	9
Seção 2: Proporção, Porcentagem, Taxa e Razão	12
Seção 3: Tabelas e Gráficos	15
Seção 4: Medidas de Tendência Central	25
Seção 5: Medidas de Variabilidade	25
Seção 6: Medidas de Posição I: Percentis	27
Seção 7: Medidas de Posição II: Escores Padronizados	29
Seção 8: Associação entre Variáveis Qualitativas	31
Seção 9: Associação entre Variáveis Quantitativas	33
Seção 10: Introdução à Amostragem	37
Seção 11: Medidas de Desigualdade	39

Segunda Parte: Solução dos Exercícios

Seção 1: Abusos da Estatística	43
Seção 2: Proporção, Porcentagem, Taxa e Razão	47
Seção 3: Tabelas e Gráficos	51
Seção 4: Medidas de Tendência Central	59
Seção 5: Medidas de Variabilidade	61
Seção 6: Medidas de Posição I: Percentis	63
Seção 7: Medidas de Posição II: Escores Padronizados	65
Seção 8: Associação entre Variáveis Qualitativas	68
Seção 9: Associação entre Variáveis Quantitativas	70
Seção 10: Introdução à Amostragem	75
Seção 11: Medidas de Desigualdade	77

Referências Bibliográficas **81**

Seção 1: Abusos da Estatística

Exercício 1: Uma funcionária tem um salário anual de R\$10.000,00, mas é informada que terá uma redução de 10% em seu salário em virtude do declínio dos lucros da companhia. Entretanto, é informada que, no próximo ano, terá um aumento de 10%. Ela aceita, acreditando que a situação não se afigura tão ruim, pois a redução inicial de 10% lhe parece ser compensada pelo aumento posterior de 10%.

- Qual será renda anual da funcionária após a redução de 10% ?
- No próximo ano, qual será a renda anual da funcionária após o aumento de 10% ?
- A redução inicial de 10% seguida do aumento posterior de 10% restitue à funcionária o salário original de R\$10.000,00 ?
- Qual deve ser, em porcentagem da sua renda, este aumento posterior para que a funcionária volte a ter uma renda de R\$10.000,00 ?

Exercício 2: Um editorial do *New York Times* criticou um anúncio que alegava que determinado anti-séptico bucal “reduz em mais de 300% as placas nos dentes”.

- Removendo-se 100% de uma quantidade, quanto resta ?
- É correto dizer que houve uma redução de mais de 300% de uma quantidade ?

Exercício 3: Em um suplemento de propaganda inserido no *Time*, os aumentos das despesas com o combate à poluição foram ilustrados em um gráfico como o que aparece a seguir (à esquerda). O que está errado com o gráfico ? Desenhe o gráfico correto.

Figura 1.3.1: Gráfico apresentado no Time.

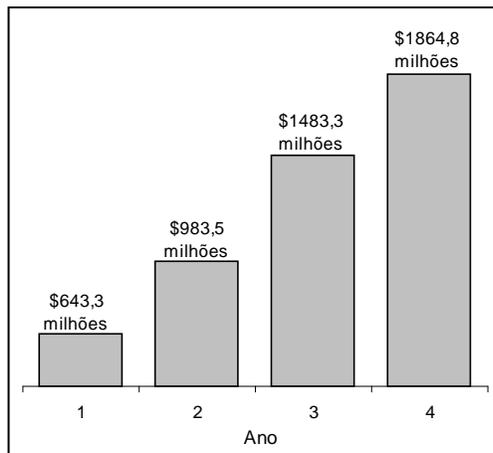
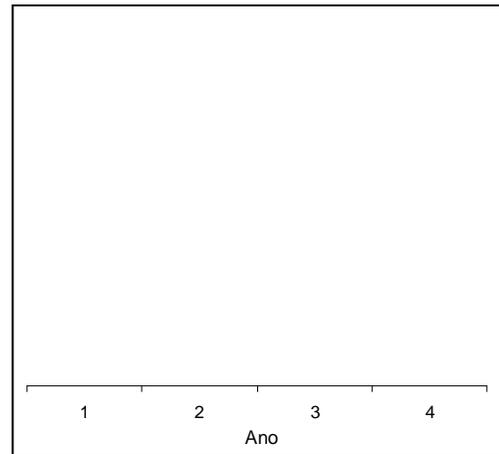


Figura 1.3.2: Desenhe o gráfico correto.



Exercício 4: Um jornalista realizou uma pesquisa solicitando a resposta dos leitores de um jornal de uma grande cidade à seguinte pergunta:

“Você apóia o desenvolvimento de armas atômicas que poderiam matar milhões de pessoas inocentes?”

Relata-se que 20 leitores ligaram para a redação respondendo à pergunta. A manchete de uma edição posterior do jornal dizia:

87% dos cidadãos não querem armas atômicas.

Identifique cinco grandes falhas neste fato.

Exercício 5: Pesquisadores do censo, ao tabular a idade das pessoas recenseadas, encontraram mais pessoas com 50 anos do que com 49 ou 51 anos; mais pessoas com 60 anos do que com 59 ou 61 anos. Como você explica este fato ?

Exercício 6: A pergunta a seguir, feita em uma pesquisa, teve repercussão quando as respostas sugeriram que cerca de 22% dos americanos achavam que o holocausto pode não ter existido (uma porcentagem considerada alta demais):

"Acha possível ou impossível que a exterminação de judeus pelos nazistas nunca tenha ocorrido?"

Uma pesquisa posterior revelou que os entrevistados provavelmente sentiram-se confusos pela dupla negativa dessa frase. Nesta pesquisa, a frase foi reformulada:

"Acha possível que a exterminação de judeus pelos nazistas jamais ocorreu ou está certo de que realmente aconteceu?"

- a) Aponte os problemas da primeira pergunta.
- b) A segunda pergunta parece substancialmente menos confusa ? Quais problemas você pode apontar nela ?
- c) Reformule a pergunta de modo que ela se apresente mais clara, mais curta e neutra.

Exercício 7: Leia o quadrinho abaixo e comente sobre o tipo de abuso da Estatística que foi cometido nesta situação.

Figura 1.7.1:



Exercício 8: Uma propaganda na TV afirma que o cursinho pré-vestibular Gama é o melhor da cidade, pois seus alunos ficaram com 48% das vagas de uma universidade. Analisando os dados na Tabela 1.8.1 abaixo, explique por que você concorda ou discorda da afirmação da propaganda.

Tabela 1.8.1: Resultados de aprovação e reprovação no vestibular de 50 mil candidatos segundo cursinho pré-vestibular freqüentado.

Cursinho	Aprovados	Reprovados	Candidatos
Alpha	1600	2400	4000
Beta	400	100	500
Gama	2400	3600	6000
Nenhum	600	38900	39500
Total	5000	45000	50000

Exercício 9: Leia o seguinte trecho de uma matéria da revista *Veja* (15/04/1998, p. 86) sobre experimentos com a droga tamoxifeno, que foi testada para verificar seus efeitos na redução do risco de vários tipos de câncer, dentre eles o câncer de mama:

“Ao longo de seis anos, os cientistas acompanharam 13.388 pacientes. Todas eram vítimas potenciais da doença — ou pelo histórico familiar ou pela idade avançada. A metade da amostra foi tratada com um comprimido de tamoxifeno por dia. A outra metade, com pílulas sem nenhum efeito terapêutico. Os resultados foram surpreendentes. Entre as mulheres tratadas com a droga, 85 desenvolveram o câncer. Sem a droga, foram 154 — 45% mais.”

Explique por que os resultados foram muito mais “surpreendentes” do que a revista sugere com a comparação dos números apresentados.

Exercício 10: Um jornalista de uma revista deseja representar, através de um pictográfico, o aumento de três vezes no número de vacas leiteiras de um estado entre os anos de 1950 e 2000. Foram colocadas duas opções para o jornalista escolher, mostradas a seguir. Ajude-o nesta decisão, levando em conta a veracidade e clareza de informação que se deseja da representação gráfica.

Figura 1.10.1: Representação (I).

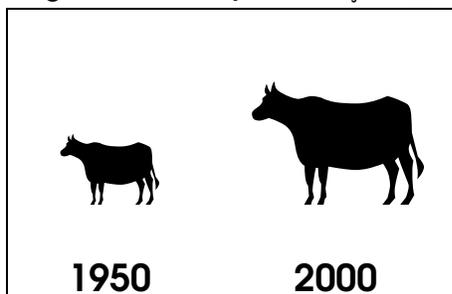
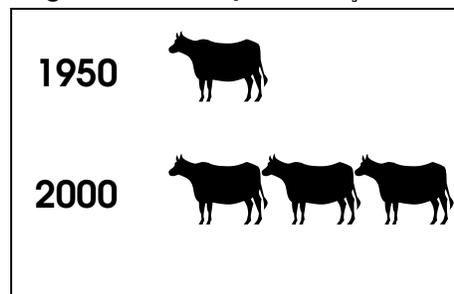


Figura 1.10.2: Representação (II).



Seção 2: Proporção, Porcentagem, Taxa e Razão

Exercício 1: Considere o Quadro 2.1.1 a seguir, que mostra a distribuição típica do peso entre os órgãos em um homem com 70 quilogramas.

Quadro 2.1.1: Distribuição do peso corporal de homem com 70 Kg.

Órgão	Peso (Kg)	Peso (%)
Músculos	30,0	42,9%
Esqueleto	10,0	
Sangue	5,0	
Gastrointestinal	2,0	
Fígado	1,7	
Cérebro	1,5	
Pulmões	1,0	
Outros	18,8	
Total	70,0	100%

- Complete o Quadro 2.1.1, calculando a participação relativa (%) de cada órgão no peso total.
- Se 40% do peso do sangue é constituído por células, qual é o peso em grams das células do sangue ?
- Qual a porcentagem do peso total do corpo que é devida às células do sangue ?
- Se o peso do esqueleto é representado por x quilos, como é representado o peso total do corpo?

Exercício 2: O Quadro 2.2.1 a seguir mostra a distribuição da população economicamente ativa ocupada de uma cidade nos diversos setores de atividade econômica no mês de março de 1990 e 2000.

Quadro 2.2.1: Distribuição (%) da PEA ocupada no mês de março em 1990 e 2000.

Mês/Ano	Indústria	Construção Civil	Comércio	Serviços	Serviços Domésticos	Outros
Março/1990	18	10	18	38	14	2
Março/2000	16	8	16	49	10	1

- Qual dos setores apresentou o maior crescimento entre os ocupados?
- Sabendo-se que a população ocupada em Março/2000 era de 1.453.000, quantas pessoas estavam trabalhando no setor de Serviços, aproximadamente?
- A categoria Outros teve um decréscimo percentual de 50% ao passar de 2% para 1%. Qual foi o decréscimo percentual da categoria Construção Civil?
- Dente os desempregados em Março/2000, 11% trabalhavam na Construção Civil. Se todos fossem recontratados na Construção Civil, qual seria a nova porcentagem desta categoria?
- Se o número de ocupados na indústria em Março/1999 era 196.740 pessoas, então qual o número total de ocupados?

Exercício 3: No Vestibular 2001 da UFMG, inscreveram-se 45.688 mulheres e 32.014 homens (COPEVE-UFMG).

- a) Calcule a proporção de mulheres no total de inscritos. Expresse esta proporção como número de mulheres para cada grupo de dez (cem, mil, dez mil) inscritos.
- b) Calcule a razão de sexos (mulheres/homens) dentre os inscritos. Expresse esta razão como número de mulheres para cada grupo de dez (cem, mil, dez mil) homens.

Exercício 4: O quadro 2.4.1 a seguir mostra os valores das alíquotas de desconto do Imposto de Renda na Fonte em Rendimentos Tributáveis de Pessoas Físicas (Receita Federal - Ministério da Fazenda). O valor do imposto sobre o rendimento é calculado como:

soma das alíquotas aplicadas à parcela do rendimento em cada nível de rendimento.

Quadro 2.4.1:

Parcela do Rendimento	Desconto
Até 900 reais	0%
900 a 1800 reais	15%
Acima de 1800 reais	27,5%

Por exemplo, para um rendimento de:

- 800 reais: 0% de 800 = 0 reais
- 1200 reais, o imposto = (0% de 900) + (15% de (1200 – 900)) = (0% de 900) + (15% de 300) = 45reais
- 2500 reais, o imposto = (0% de 900) + (15% de (1800 – 900)) + (27,5% de (2500-1800))
= (0% de 900) + (15% de 900) + (27,5% de 700) = 135 + 192,5 = 327,50 reais.

a) Calcule o valor do imposto para os seguintes rendimentos:

- i) Rendimento = 875 reais
- ii) Rendimento = 1400 reais
- iii) Rendimento = 3000 reais

Entretanto, a forma usual, mais fácil, de se fazer o cálculo do imposto é:

alíquota sobre rendimento – desconto,

onde as alíquotas e desconto referentes a cada classe de valor do rendimento são:

Quadro 2.4.2:

Valor do Rendimento	Alíquota	Desconto
0 a 900 reais	0%	0 reais
900 a 1800 reais	15%	135 reais
Acima de 1800 reais	27,5%	360 reais

Por exemplo, para um rendimento de:

- 800 reais, o imposto = (0% de 800) – 0 = 0 reais;
- 1200 reais, o imposto = (15% de 1200) – 135 = 210 – 135 = 45 reais;
- 2500 reais, o imposto = (27,5% de 2500) – 360 = 687,50 – 360 = 327,50 reais.

b) Verifique, usando como exemplos os valores de rendimentos do item (a), que esta forma de cálculo equivale à anterior.

Exercício 5: O Quadro 2.5.1 a seguir mostra dados sobre inscrição e pedidos de isenção da taxa de inscrição no Vestibular UFMG de 1997 a 2002 (COPEVE-UFMG).

Quadro 2.5.1:

Ano	1997	1998	1999	2000	2001	2002
Pedidos de Isenção*	5284	11056	12286	22189	30425	54234
Deferidos	4322	7985	6543	11303	18770	31362
Indeferidos	962	3071	5743	10886	11655	22872
Inscritos no Vestibular	45994	51073	61749	77702	78889	n.d.
Isentos*	3813	6884	5759	10349	16667	n.d.
Não Isentos*	42181	44189	55990	67353	62222	n.d.
Aprovados no Vestibular	3801	3877	4017	4177	4362	n.d.
Isentos*	200	258	211	216	381	n.d.
Não Isentos*	3601	3619	3806	3961	3961	n.d.

*Da taxa de inscrição. n.d.: informação não disponível nesta data.

Analise cada afirmação a seguir e comente sobre sua veracidade baseado nos dados do Quadro 2.5.1:

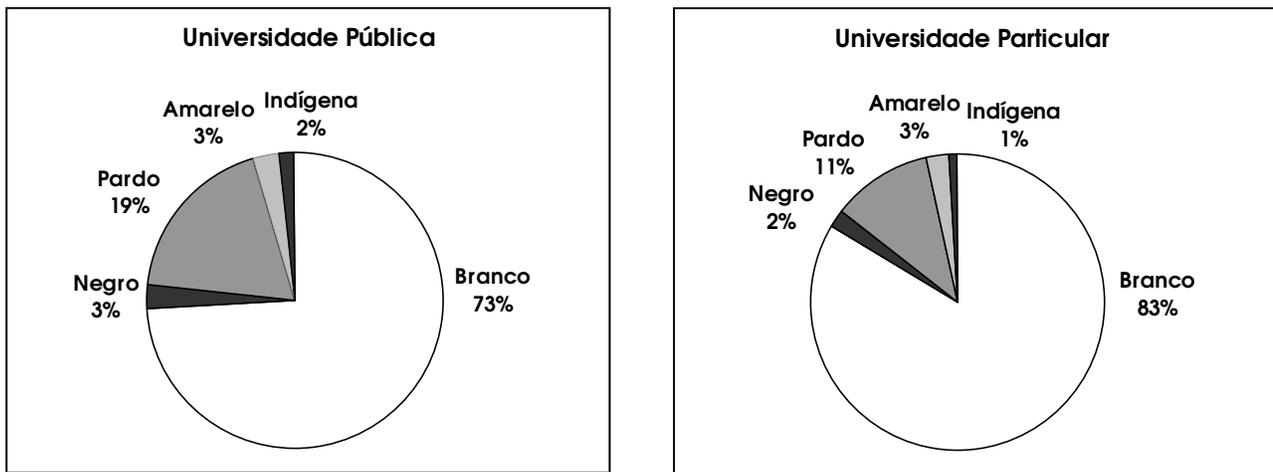
- i. Ao longo dos últimos seis anos, a UFMG vem dando mais apoio aos estudantes carentes que querem prestar o Vestibular. Prova disso é o número de pedidos de isenção de taxa de inscrição deferidos, que aumentou de 4322 em 1997 para 31362 no ano 2000 – um crescimento de mais de 600%.
- ii. Nos dois últimos vestibulares, a UFMG não conseguiu acompanhar a demanda por isenção da taxa de inscrição: do ano 2001 para 2002, o número de pedidos de isenção recebidos aumentou 78%, enquanto que o número de pedidos deferidos aumentou 67%.
- iii. Nos dois últimos vestibulares, a UFMG não conseguiu acompanhar a demanda por isenção da taxa de inscrição: em 2001, a porcentagem de pedidos de isenção deferidos foi de 61,7%, enquanto em 2002 foi de 57,8%.
- iv. A aprovação no Vestibular da UFMG é menor para os candidatos que conseguiram isenção da taxa de inscrição. No ano 2001, por exemplo, apenas 2,3% dos candidatos isentos foram aprovados, enquanto que, para os candidatos não isentos da taxa, a porcentagem de aprovação foi de 6,4%.
- v. A aprovação no Vestibular da UFMG é menor para os candidatos que conseguiram isenção da taxa de inscrição. No ano 2001, por exemplo, os candidatos isentos representaram 21,1% do total de inscritos, mas apenas 8,7% do total de aprovados.
- vi. A aprovação no Vestibular da UFMG é menor para os candidatos que conseguiram isenção da taxa de inscrição. No ano 2001, por exemplo, os candidatos isentos representaram apenas 8,7% do total de aprovados.

Seção 3: Tabelas e Gráficos

Exercício 1: Uma análise de descarrilamentos de trens (dados da *Federal Railroad Administration*) mostrou que, dos 50 descarrilamentos registrados em um período: 23 foram causados por más condições da linha, 12 foram atribuídos a erro humano, 9 foram devidos às falhas no equipamento e 6 tiveram outras causas. Construa uma tabela de freqüências e um gráfico em setores para representar estes dados.

Exercício 2: O jornal Folha SP (1º/09/2001) fez dois gráficos de setores como os mostrados abaixo para ilustrar a matéria intitulada “Universidades públicas são menos elitistas”, que relata os resultados da pesquisa do INEP sobre a distribuição racial dos formandos nos cursos superiores avaliados pelo Provão 2001. O objetivo dos gráficos é comparar as porcentagens de cada raça nos dois gráficos, concluindo que “a universidade pública no Brasil é menos elitista, do ponto de vista racial, do que a particular” (sic).

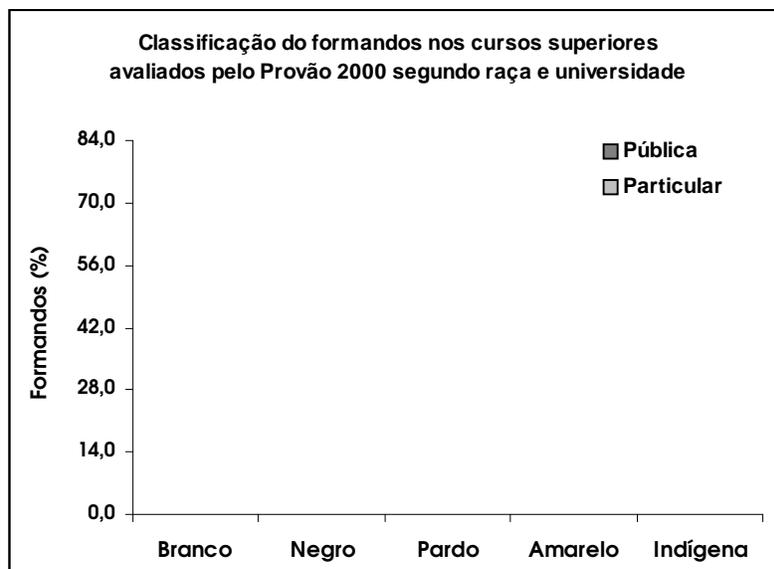
Figura 3.2.1: Gráficos em setores similares aos mostrados no jornal*.



*Os gráficos originais não puderam ser reproduzidos aqui por serem coloridos. As porcentagens foram arredondadas.

- a) Por que os gráficos de setores não cumprem bem este objetivo ?
- b) Complete o gráfico em barras da Figura 3.3.2, colocando lado-a-lado as porcentagens de cada raça nas universidades pública e particular. Por que neste gráfico é mais fácil visualizar as diferenças desejadas ?

Figura 3.2.2: Gráfico em Barras Verticais.



Exercício 3: A Fundação SEADE divulgou (*Folha Online*, 29/08/2001) estatísticas de desemprego para homens e mulheres, negros e não negros. Segundo os pesquisadores, a comparação destas as taxas de desemprego funciona como um “termômetro” da discriminação racial e/ou sexual. Eis as taxas de desemprego naquela época:

Entre negros: 21,7%; Entre não-negros: 15,7%.
 Entre mulheres: 23,9%; Entre homens: 18,1%.
 Entre mulheres negras 25,1%; Entre mulheres não-negras 18,9%.
 Entre homens negros 19%; Entre homens não-negros 13,2%

Faça um gráfico de barras verticais para mostrar:

- a) O efeito da raça na taxa de desemprego.
- b) O efeito do sexo na taxa de desemprego.
- c) O efeito simultâneo da raça e do sexo na taxa de desemprego.
- d) O efeito da raça na taxa de desemprego dentro de cada sexo.
- e) O efeito sexo na taxa de desemprego dentro de cada raça.

Exercício 4: Considere os seguintes dados sobre as importações e exportações brasileiras (em bilhões de dólares), mês a mês, durante o ano de 1998.

Importações:
 janeiro, 4577 ; fevereiro, 3799 ; março, 5038 ; abril, 4779 ; maio, 4913 ;
 junho, 4844; julho, 5329 ; agosto, 4634 ; setembro, 5338 ;
 outubro, 5039 ; novembro, 4709 ; dezembro, 4538.

Exportações:
 janeiro, 3914 ; fevereiro, 3714 ; março, 4273 ; abril, 4572 ; maio, 4609 ;
 junho, 4886 ; julho, 4970 ; agosto, 3985 ; setembro, 4537 ;
 outubro, 4014 ; novembro, 3702 ; dezembro, 3944.

Represente estas informações em um gráfico de linhas para descrever e comparar a evolução temporal das importações e exportações brasileiras.

Exercício 5: Considere as seguintes informações (dados do IPEA, publicados no jornal *Estado de Minas* em 08/07/2001):

Na população brasileira em 1999, 45,33% das pessoas são negros e 54,02% são brancos; os negros têm participação em 63,63% da população pobre e os brancos, 35,95%; na população indigente, 68,85% são negros e 30,73% são brancos.

Faça um gráfico para representar estas estatísticas e comente sobre a diferença racial na distribuição da renda baseado nestes números. Por que é necessária a informação sobre a composição racial de toda a população neste caso ?

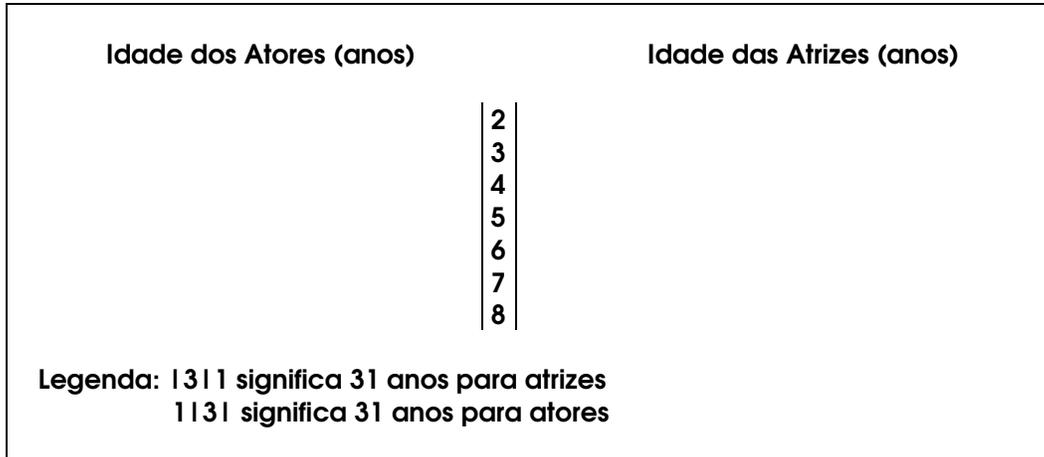
Exercício 6: No artigo “Idades dos Atores e Atrizes Ganhadores do Oscar” (revista *Mathematics Teacher*, de Richard Brown e Gretchen Davis), utilizam-se gráficos ramo-e-folhas para comparar as idades de atores e atrizes no momento da premiação. Eis os resultados para os 34 últimos vencedores até 1999:

Quadro 3.6.1:

Atores:	32	37	36	32	51	53	33	61	35	45	55	39	76	37	42	40	32
	60	38	56	48	48	40	43	62	43	42	44	41	56	39	46	31	47
Atrizes:	50	44	35	80	26	28	41	21	61	38	49	33	74	30	33	41	31
	35	41	42	37	26	34	34	35	26	61	60	34	24	30	37	31	27

Construa um ramo-e-folhas conjugado para esses dados e compare a distribuição da variável “Idade ao ganhar o Oscar” entre atores e atrizes.

Figura 3.6.1: Ramo-e-folhas para “Idade ao ganhar o Oscar”.



Exercício 7: Os dados da Tabela 3.7.1 a seguir foram obtidos numa faculdade e são referentes à idade dos carros de funcionários, professores e alunos:

Tabela 3.7.1:

Idade (anos)	Estudantes		Funcionários e Professores	
	Freqüência Absoluta	Freqüência Relativa (%)	Freqüência Absoluta	Freqüência Relativa (%)
01-12	23		30	
31-15	33		47	
61-18	63		36	
91-111	68		30	
121-114	19		8	
151-117	10		0	
181-120	1		0	
211-123	0		1	
Total	217		152	

- a) Complete a Tabela 3.7.1 e construa um histograma para cada grupo. Lembre-se de usar as freqüências relativas (pois os grupos têm amostras de tamanhos diferentes) e a mesma escala (no eixo vertical), para que os gráficos fiquem comparáveis.
- b) Analisando os histogramas, quais são as diferenças perceptíveis entre os dois grupos quanto a variável idade dos carros ?

Exercício 8: Em um estudo sobre acidentes fatais com veículos motorizados no estado de Nova York (*New York State Department of Motor Vehicles*), classificaram-se as 2068 ocorrências de acidentes fatais de acordo com a hora do dia. Os resultados foram:

Quadro 3.8.1:

Horário:	01-2	21-4	41-6	61-8	81-10	101-12	121-14	141-16	161-18	181-20	201-22	221-24
Contagem:	194	149	100	131	119	160	152	221	230	211	223	178

- a) Complete o gráfico circular a seguir.

Figura 3.8.1: Gráfico Circular.

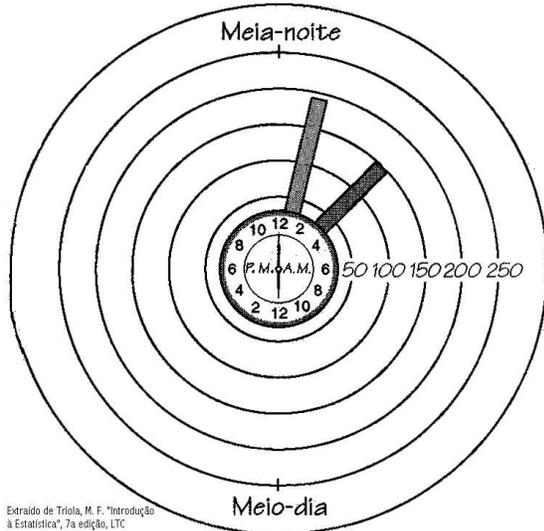
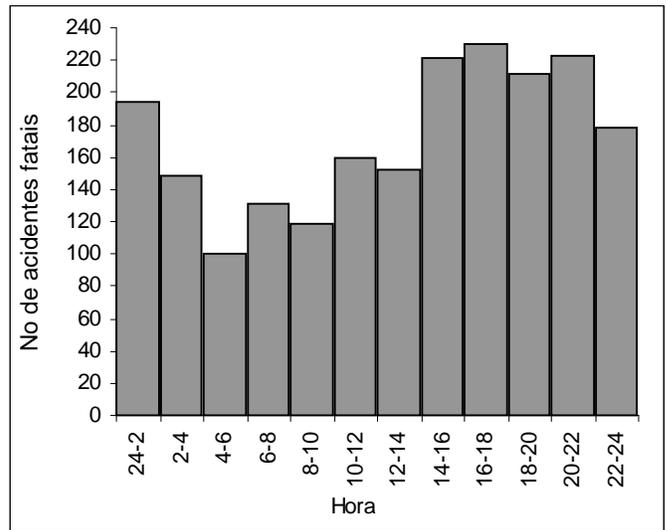


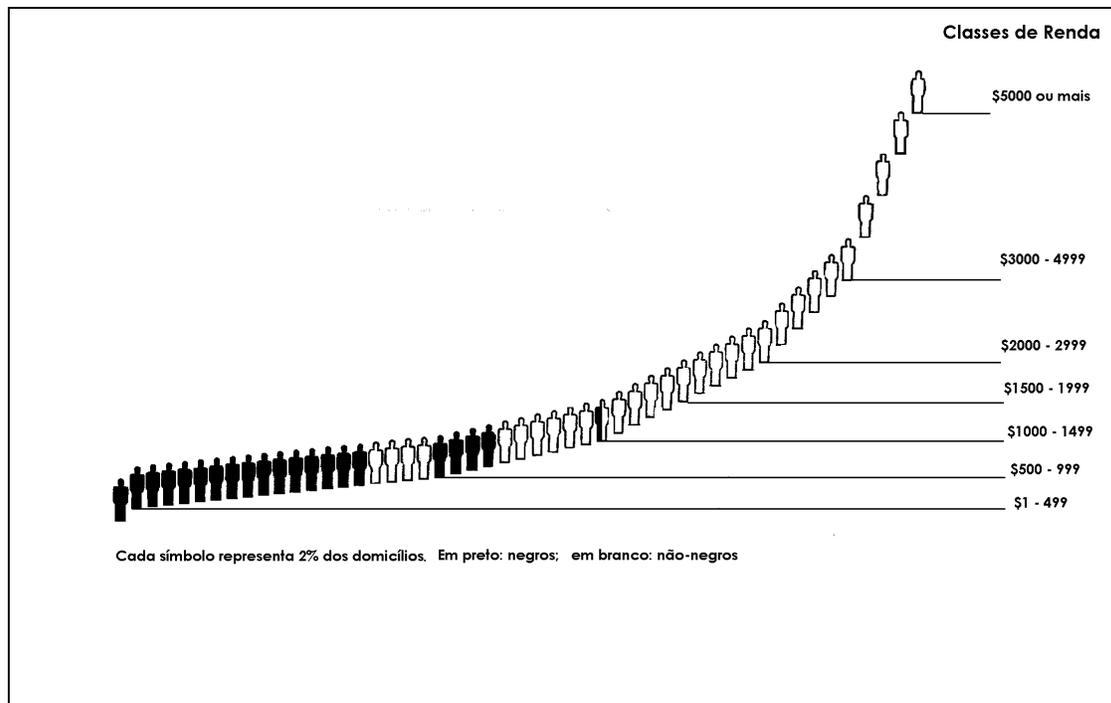
Figura 3.8.2: Histograma.



- b) Qual dos dois gráficos ilustra melhor os dados, o gráfico circular ou o histograma ? Por quê ?
- c) Como o período de 4 às 6 horas da manhã é o que acusa menor número de acidentes fatais, podemos concluir que esse período é o mais seguro para dirigir ? Por que ?

Exercício 9: Considere a Figura 3.9.1 a seguir, que mostra a distribuição de renda de negros e brancos em Columbia, Carolina do Sul, EUA, em 1933 (Zeisel, 1985).

Figura 3.9.1: Distribuição de renda de negros e brancos em Columbia, Carolina do Sul, EUA, em 1933.



Retirado de Zeisel (1985).

- a) Quais são as classes de renda consideradas nesse gráfico (Figura 3.9.1) ? Todas têm a mesma amplitude ? Como foram representadas no gráfico estas diferenças de amplitude intraclasse ?
- b) O que cada símbolo (♂ ou ♀) representa ?
- c) O que você pode dizer sobre a distribuição (forma, tendência central, variabilidade) da variável renda familiar nesta cidade ?
- d) O que você pode dizer a distribuição de renda entre raças (negros/não negros) nesta cidade ?
- e) A partir da Figura 3.9.1, complete a Tabela 3.9.1 abaixo:

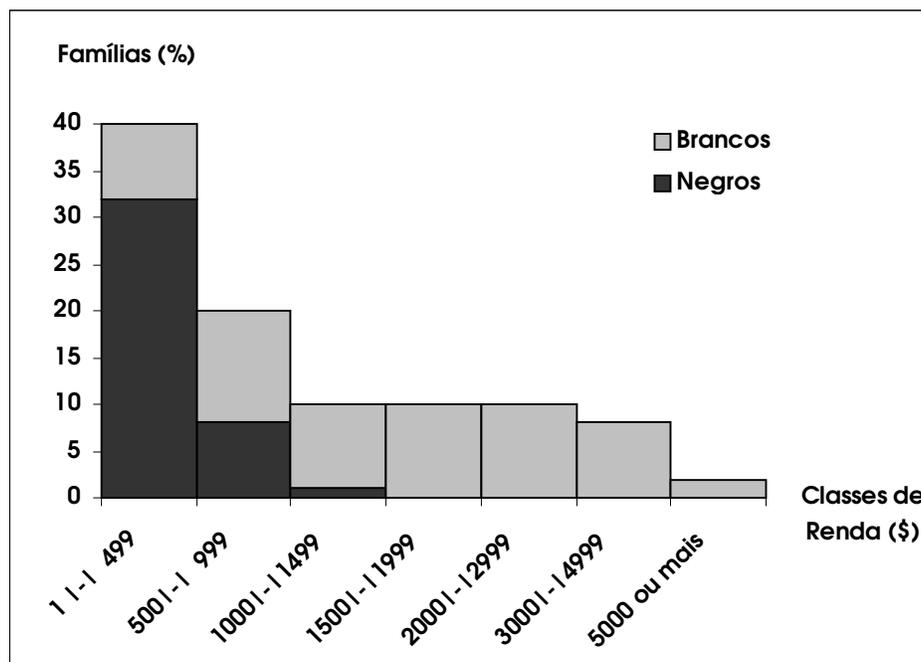
Tabela 3.9.1:

Renda (US\$)	Número de Símbolos*			Frequência Relativa (%)		
	Negros	Branco	Total	Negros	Branco	Total
1 - 499	16	4	20	32%	8%	40%
500 - 999						
1000 - 1499	0,5	4,5	5	1%	9%	10%
1500 - 1999						
2000 - 2999						
3000 - 4999						
5000 ou mais						
Total	20,5	29,5	50	41%	59%	100%

*Convenção: um símbolo faz parte de uma classe quando sua base ("pés") faz parte desta classe.

- f) O gráfico em barras verticais da Figura 3.9.2 a seguir foi construído a partir das frequências relativas da Tabela 3.9.1. Compare esta figura com a Figura 3.9.1, apontando as suas vantagens e desvantagens na representação das características da variável renda domiciliar nesta cidade que você descreveu nos itens (c) e (d).

Figura 3.9.2:



Exercício 10: O Quadro 3.10.1, na página seguinte, apresenta os dados de 200 réus condenados à prisão por roubo ou furto em uma cidade, classificados quanto ao sexo, à severidade da sentença recebida na atual condenação e o número de condenações anteriores por outros crimes. A partir deste quadro, complete as tabelas a seguir com as contagens correspondentes.

Tabela 3.10.1: Sexo X Severidade da Sentença.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino			
Feminino			
Total			

Tabela 3.10.2: Sexo X Número de Condenações Anteriores.

Sexo	Condenações Anteriores			Total
	Nenhuma	Uma ou duas	Três ou mais	
Masculino				
Feminino				
Total				

Tabela 3.10.3: Severidade da Sentença X Número de Condenações Anteriores.

Severidade Sentença	Condenações Anteriores			Total
	Nenhuma	Uma ou duas	Três ou mais	
Moderada				
Severa				
Total				

Tabela 3.10.4: Sexo X Severidade da Sentença segundo Número de Condenações Anteriores.

a) Nenhuma condenação anterior.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino			
Feminino			
Total			

b) Uma ou duas condenações anteriores.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino			
Feminino			
Total			

c) Três ou mais condenações anteriores.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino			
Feminino			
Total			

Quadro 3.10.1: Classificação de 200 réus condenados à prisão por roubo ou furto, quanto ao sexo, a severidade da sentença recebida e o número de condenações anteriores por outros crimes.

Réu	Sexo	Severidade da Sentença	Condenações Anteriores
1	Feminino	Moderada	Nenhuma
2	Feminino	Moderada	Nenhuma
3	Feminino	Moderada	Nenhuma
4	Feminino	Moderada	Nenhuma
5	Feminino	Moderada	Nenhuma
6	Feminino	Moderada	Nenhuma
7	Feminino	Moderada	Nenhuma
8	Feminino	Moderada	Nenhuma
9	Feminino	Moderada	Nenhuma
10	Feminino	Moderada	Nenhuma
11	Feminino	Moderada	Nenhuma
12	Feminino	Moderada	Nenhuma
13	Feminino	Moderada	Nenhuma
14	Feminino	Moderada	Nenhuma
15	Feminino	Moderada	Nenhuma
16	Feminino	Moderada	Nenhuma
17	Feminino	Moderada	Nenhuma
18	Feminino	Moderada	Nenhuma
19	Feminino	Moderada	Nenhuma
20	Feminino	Moderada	Nenhuma
21	Feminino	Moderada	Nenhuma
22	Feminino	Moderada	Nenhuma
23	Feminino	Moderada	Nenhuma
24	Feminino	Moderada	Nenhuma
25	Feminino	Moderada	Nenhuma
26	Feminino	Moderada	Nenhuma
27	Feminino	Moderada	Nenhuma
28	Feminino	Moderada	Nenhuma
29	Feminino	Moderada	Nenhuma
30	Feminino	Moderada	Nenhuma
31	Feminino	Moderada	Nenhuma
32	Feminino	Moderada	Nenhuma
33	Feminino	Moderada	Nenhuma
34	Feminino	Moderada	Nenhuma
35	Feminino	Moderada	Nenhuma
36	Feminino	Moderada	Nenhuma
37	Feminino	Moderada	Nenhuma
38	Feminino	Moderada	Nenhuma
39	Feminino	Moderada	Nenhuma
40	Feminino	Moderada	Nenhuma
41	Feminino	Moderada	Nenhuma
42	Feminino	Moderada	Nenhuma
43	Feminino	Moderada	Nenhuma
44	Feminino	Moderada	Nenhuma
45	Feminino	Moderada	Nenhuma
46	Feminino	Moderada	Nenhuma
47	Feminino	Moderada	Nenhuma
48	Feminino	Moderada	Nenhuma
49	Feminino	Moderada	Nenhuma
50	Feminino	Moderada	Nenhuma

Réu	Sexo	Severidade da Sentença	Condenações Anteriores
51	Feminino	Severa	Nenhuma
52	Feminino	Severa	Nenhuma
53	Feminino	Severa	Nenhuma
54	Feminino	Severa	Nenhuma
55	Feminino	Severa	Nenhuma
56	Feminino	Severa	Nenhuma
57	Feminino	Severa	Nenhuma
58	Feminino	Severa	Nenhuma
59	Feminino	Severa	Nenhuma
60	Feminino	Severa	Nenhuma
61	Feminino	Severa	Nenhuma
62	Feminino	Severa	Nenhuma
63	Feminino	Severa	Nenhuma
64	Feminino	Severa	Nenhuma
65	Feminino	Severa	Nenhuma
66	Feminino	Severa	Nenhuma
67	Feminino	Severa	Nenhuma
68	Feminino	Severa	Nenhuma
69	Feminino	Severa	Nenhuma
70	Feminino	Severa	Nenhuma
71	Feminino	Moderada	Uma ou duas
72	Feminino	Moderada	Uma ou duas
73	Feminino	Moderada	Uma ou duas
74	Feminino	Moderada	Uma ou duas
75	Feminino	Moderada	Uma ou duas
76	Feminino	Moderada	Uma ou duas
77	Feminino	Moderada	Uma ou duas
78	Feminino	Moderada	Uma ou duas
79	Feminino	Moderada	Uma ou duas
80	Feminino	Moderada	Uma ou duas
81	Feminino	Severa	Uma ou duas
82	Feminino	Severa	Uma ou duas
83	Feminino	Severa	Uma ou duas
84	Feminino	Severa	Uma ou duas
85	Feminino	Severa	Uma ou duas
86	Feminino	Severa	Uma ou duas
87	Feminino	Severa	Uma ou duas
88	Feminino	Severa	Uma ou duas
89	Feminino	Severa	Uma ou duas
90	Feminino	Severa	Uma ou duas
91	Feminino	Severa	Três ou mais
92	Feminino	Severa	Três ou mais
93	Feminino	Severa	Três ou mais
94	Feminino	Severa	Três ou mais
95	Feminino	Severa	Três ou mais
96	Feminino	Severa	Três ou mais
97	Feminino	Severa	Três ou mais
98	Feminino	Severa	Três ou mais
99	Feminino	Severa	Três ou mais
100	Feminino	Severa	Três ou mais

Continua...

Continuação: Quadro 3.10.1.

Réu	Sexo	Severidade da Sentença	Condenações Anteriores
101	Masculino	Moderada	Nenhuma
102	Masculino	Moderada	Nenhuma
103	Masculino	Moderada	Nenhuma
104	Masculino	Moderada	Nenhuma
105	Masculino	Moderada	Nenhuma
106	Masculino	Moderada	Nenhuma
107	Masculino	Moderada	Nenhuma
108	Masculino	Moderada	Nenhuma
109	Masculino	Moderada	Nenhuma
110	Masculino	Moderada	Nenhuma
111	Masculino	Moderada	Uma ou duas
112	Masculino	Moderada	Uma ou duas
113	Masculino	Moderada	Uma ou duas
114	Masculino	Moderada	Uma ou duas
115	Masculino	Moderada	Uma ou duas
116	Masculino	Moderada	Uma ou duas
117	Masculino	Moderada	Uma ou duas
118	Masculino	Moderada	Uma ou duas
119	Masculino	Moderada	Uma ou duas
120	Masculino	Moderada	Uma ou duas
121	Masculino	Moderada	Uma ou duas
122	Masculino	Moderada	Uma ou duas
123	Masculino	Severa	Uma ou duas
124	Masculino	Severa	Uma ou duas
125	Masculino	Severa	Uma ou duas
126	Masculino	Severa	Uma ou duas
127	Masculino	Severa	Uma ou duas
128	Masculino	Severa	Uma ou duas
129	Masculino	Severa	Uma ou duas
130	Masculino	Severa	Uma ou duas
131	Masculino	Moderada	Três ou mais
132	Masculino	Moderada	Três ou mais
133	Masculino	Moderada	Três ou mais
134	Masculino	Moderada	Três ou mais
135	Masculino	Moderada	Três ou mais
136	Masculino	Moderada	Três ou mais
137	Masculino	Moderada	Três ou mais
138	Masculino	Moderada	Três ou mais
139	Masculino	Moderada	Três ou mais
140	Masculino	Moderada	Três ou mais
141	Masculino	Moderada	Três ou mais
142	Masculino	Moderada	Três ou mais
143	Masculino	Moderada	Três ou mais
144	Masculino	Moderada	Três ou mais
145	Masculino	Moderada	Três ou mais
146	Masculino	Moderada	Três ou mais
147	Masculino	Moderada	Três ou mais
148	Masculino	Moderada	Três ou mais
149	Masculino	Severa	Três ou mais
150	Masculino	Severa	Três ou mais

Réu	Sexo	Severidade da Sentença	Condenações Anteriores
151	Masculino	Severa	Três ou mais
152	Masculino	Severa	Três ou mais
153	Masculino	Severa	Três ou mais
154	Masculino	Severa	Três ou mais
155	Masculino	Severa	Três ou mais
156	Masculino	Severa	Três ou mais
157	Masculino	Severa	Três ou mais
158	Masculino	Severa	Três ou mais
159	Masculino	Severa	Três ou mais
160	Masculino	Severa	Três ou mais
161	Masculino	Severa	Três ou mais
162	Masculino	Severa	Três ou mais
163	Masculino	Severa	Três ou mais
164	Masculino	Severa	Três ou mais
165	Masculino	Severa	Três ou mais
166	Masculino	Severa	Três ou mais
167	Masculino	Severa	Três ou mais
168	Masculino	Severa	Três ou mais
169	Masculino	Severa	Três ou mais
170	Masculino	Severa	Três ou mais
171	Masculino	Severa	Três ou mais
172	Masculino	Severa	Três ou mais
173	Masculino	Severa	Três ou mais
174	Masculino	Severa	Três ou mais
175	Masculino	Severa	Três ou mais
176	Masculino	Severa	Três ou mais
177	Masculino	Severa	Três ou mais
178	Masculino	Severa	Três ou mais
179	Masculino	Severa	Três ou mais
180	Masculino	Severa	Três ou mais
181	Masculino	Severa	Três ou mais
182	Masculino	Severa	Três ou mais
183	Masculino	Severa	Três ou mais
184	Masculino	Severa	Três ou mais
185	Masculino	Severa	Três ou mais
186	Masculino	Severa	Três ou mais
187	Masculino	Severa	Três ou mais
188	Masculino	Severa	Três ou mais
189	Masculino	Severa	Três ou mais
190	Masculino	Severa	Três ou mais
191	Masculino	Severa	Três ou mais
192	Masculino	Severa	Três ou mais
193	Masculino	Severa	Três ou mais
194	Masculino	Severa	Três ou mais
195	Masculino	Severa	Três ou mais
196	Masculino	Severa	Três ou mais
197	Masculino	Severa	Três ou mais
198	Masculino	Severa	Três ou mais
199	Masculino	Severa	Três ou mais
200	Masculino	Severa	Três ou mais

Seção 4: Medidas de Tendência Central

Exercício 1: A seguir são mostrados os tempos de espera em fila (em minutos) de clientes em um banco onde os clientes forma uma fila única no (Banco A) e em outro banco onde os clientes entram em três filas de guichês (Banco de B). Determine a média, a mediana e a moda em cada grupo e compare os dois tipos de filas quanto à tendência central da variável "tempo de espera".

Quadro 4.1.1:

Banco A (fila única):	6,5	6,6	6,7	6,8	7,1	7,3	7,4	7,7	7,7	7,7
Banco B (fila múltipla):	4,2	5,4	5,8	6,2	6,7	7,7	7,7	8,5	9,3	10,0

Exercício 2: O quadro 4.2.1 a seguir mostra os resultados da medição da largura máxima de amostras de crânios de egípcios do sexo masculino, referentes a duas épocas diferentes: 4000 aC e 150 dC (*Ancient Races of the Thebaid*, de Thomson e Randall-Maciver):

Quadro 4.2.1:

4000 aC:	131	199	138	125	129	126	131	132	126	128	128	131
150 dC:	136	130	126	126	139	141	137	138	133	131	134	129

Determine a média, a mediana e a moda de cada em cada grupo e compare as duas épocas quanto à tendência central da variável "largura máxima de crânios de egípcios do sexo masculino".

Exercício 3: Em uma pesquisa realizada com 15 mulheres, perguntou-se a idade. Doze destas mulheres responderam informaram o valor (mostrados no ramo-e-folhas a seguir), mas três delas disseram apenas que tinham mais de 30 anos.

Figura 4.3.1: Ramo-e-Folhas para a Idade de 12 das 15 mulheres da pesquisa.

```
1 | 4
1 | 8 9
2 | 0 2 4
2 | 6 7 7
3 | 1 3
3 | 5
```

Legenda: 114 = 14 anos

Que medida de tendência central, média ou mediana, você considera mais adequada para descrever a idade típica destas 15 mulheres? Explique.

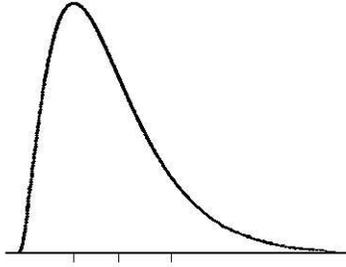
Exercício 4: Responda se cada uma das afirmativas a seguir é verdadeira ou falsa. Se a afirmativa for falsa, corrija a palavra sublinhada para que se torne verdadeira.

- Metade dos valores de um variável quantitativa são sempre menores que a média.
- Quando a variável quantitativa tem distribuição unimodal e simétrica, metade de seus valores é menor que a média.
- A mediana não é uma boa medida de tendência central para uma variável quantitativa com distribuição unimodal muito assimétrica, pois esta medida é muito influenciada por valores extremos.
- Quando a variável quantitativa tem distribuição bimodal, a mediana é a medida de tendência central mais adequada.

e) Quando a variável quantitativa tem distribuição unimodal e simétrica, a posição relativa das medidas e tendência central é : média < mediana < moda.

Exercício 5: Nos esboços de distribuições de freqüências mostrados a seguir, marque as posições relativas da moda, média e mediana de acordo com a forma da distribuição.

Figura 4.5.1:



Seção 5: Medidas de Variabilidade

Exercício 1: Dois grupos diferentes de uma turma de estatística fazem o mesmo teste-surpresa, com as notas relacionadas a seguir.

Quadro 5.1.1:

Grupo 1:	1	20	20	20	20	20	20	20	20	20	20
Grupo 2:	2	3	4	5	6	14	15	16	17	18	19

- Encontre a amplitude total e o desvio padrão para cada grupo.
- Que conclusões sobre a variação nos dois grupos os valores da amplitude total sugerem?
- Por que a amplitude total é enganosa neste caso?
- Que conclusões sobre a variação nos dois grupos o desvio-padrão sugere?

Exercício 2: A seguir são mostrados os tempos de espera em fila (em minutos) de clientes em um banco onde os clientes forma uma fila única no (Banco A) e em outro banco onde os clientes entram em três filas de guichês (Banco de B):

Quadro 5.2.1:

Banco A (fila única):	6,5	6,6	6,7	6,8	7,1	7,3	7,4	7,7	7,7	7,7
Banco B (fila múltipla):	4,2	5,4	5,8	6,2	6,7	7,7	7,7	8,5	9,3	10,0

Calcule a amplitude total, o desvio-padrão e o coeficiente de variação para cada uma das duas amostras. Compare os resultados dos dois tipos de filas utilizando estas medidas de dispersão que você acabou de calcular e as medidas de tendência central mostradas a seguir.

Quadro 5.2.2:

Banco (tipo de fila)	Medidas de Tendência Central			Medidas de Variabilidade		
	Média	Mediana	Moda	Amplitude Total	Desvio Padrão	Coefficiente de variação
A (fila única)	7,15	7,2	7,7			
B (fila múltipla)	7,15	7,2	7,7			

Exercício 3: Como administrador, o leitor deve comprar lâmpadas para um hospital. Escolheria as lâmpadas *Ultralight*, que têm vida média de 3000 horas e desvio padrão de 200 horas, ou as lâmpadas *Electrolyte*, com vida média de 3000 horas e desvio padrão de 250 horas ? Explique.

Exercício 4: Para comparar a precisão de dois micrômetros, um técnico estuda medidas tomadas com ambos os aparelhos. Com um, mediu repetidamente o diâmetro de uma pequena esfera de rolamento; as mensurações acusaram média de 5,32 mm e desvio-padrão de 0,019 mm. Com o outro, mediu o comprimento natural de uma mola, tendo as mensurações acusado média de 6,4 cm e desvio-padrão de 0.03 cm. Supondo que o verdadeiro valor do diâmetro da esfera seja de 5,32 mm e do comprimento da mola seja de 6,4 qual dos dois aparelhos é relativamente mais preciso?

Exercício 5: O quadro 5.5.1 a seguir apresenta a média e o desvio padrão dos valores de três variáveis usadas como indicadores da qualidade de vida de uma população, medidas nas 27 unidades da federação brasileiras (*Folha de São Paulo* de 09/09/98):

Quadro 5.5.1:

Variável	Média	Desvio Padrão
Esperança de Vida (em anos)	66,97	2,39
Taxa de Alfabetização (em %)	80,85	9,62
PIB per capita ajustado (em dólares ajustados pelo poder de compra)	4846	1697

Deseja-se escolher uma dessas variáveis para se criar um *ranking* de qualidade de vida entre estas UF's. A variável escolhida deve ser, portanto, a de melhor poder de discriminação. Com as informações fornecidas no Quadro 5.5.1, podemos dizer que esta deve ser a variável de maior heterogeneidade entre as UF's. Qual deve ser a variável escolhida ? Justifique.

Seção 6: Medidas de Posição I

Percentis

Exercício 1: O ramo-e-folhas a seguir mostra os dados de renda mensal (em salários mínimos) de chefes de família em 100 domicílios de uma comunidade. As contagens entre parênteses referem-se às frequências absolutas acumuladas em cada classe, incluindo as observações daquela classe.

Figura 6.1.1: Ramo-e-folhas para Renda Mensal.

(18)	1	12555566666677889
(35)	2	01112233444467889
(50)	3	002222334567899
(64)	4	01234466667789
(73)	5	001255689
(82)	6	001244567
(90)	7	00125588
(96)	8	011249
(99)	9	117
(100)	10	8

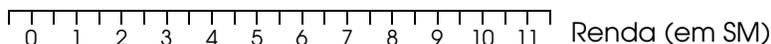
Legenda:
111 = 1,1
salários mínimos

- a) Calcule a mediana, o primeiro e o terceiro quartil a renda mensal nesta amostra. Interprete.
- b) Construa, gráfico abaixo, o Box-Plot para a renda mensal nesta amostra.

$$DQ = Q3 - Q1 = \underline{\quad} - \underline{\quad} = \underline{\quad} \qquad Q1 - 1,5 \times DQ = \underline{\quad} - \underline{\quad} = \underline{\quad}$$

$$1,5 \times DQ = 1,5 \times \underline{\quad} = \underline{\quad} \qquad Q3 + 1,5 \times DQ = \underline{\quad} - \underline{\quad} = \underline{\quad}$$

Figura 6.1.2: Box-Plot.



- c) Construa a Ogiva de Frequências Relativas Acumuladas (em %) no gráfico abaixo.

Exemplo de construção:

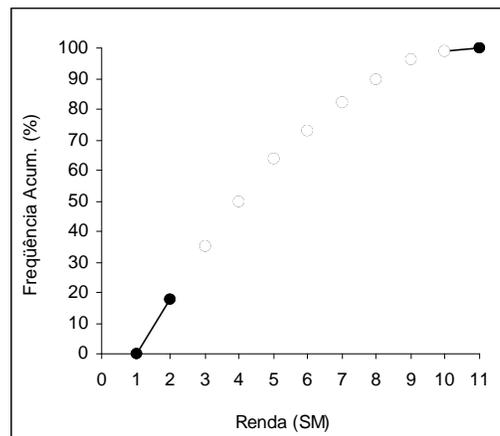
Abaixo de 1 s.m. não há nenhuma observação; portanto, o primeiro ponto do gráfico corresponde ao par (1; 0).

Abaixo de 2 s.m. temos 18 observações (18/100 - 18%); portanto, o segundo ponto do gráfico corresponde ao par (2; 18).

Faça o mesmo para 3, 4, 5, ..., 10 s.m.

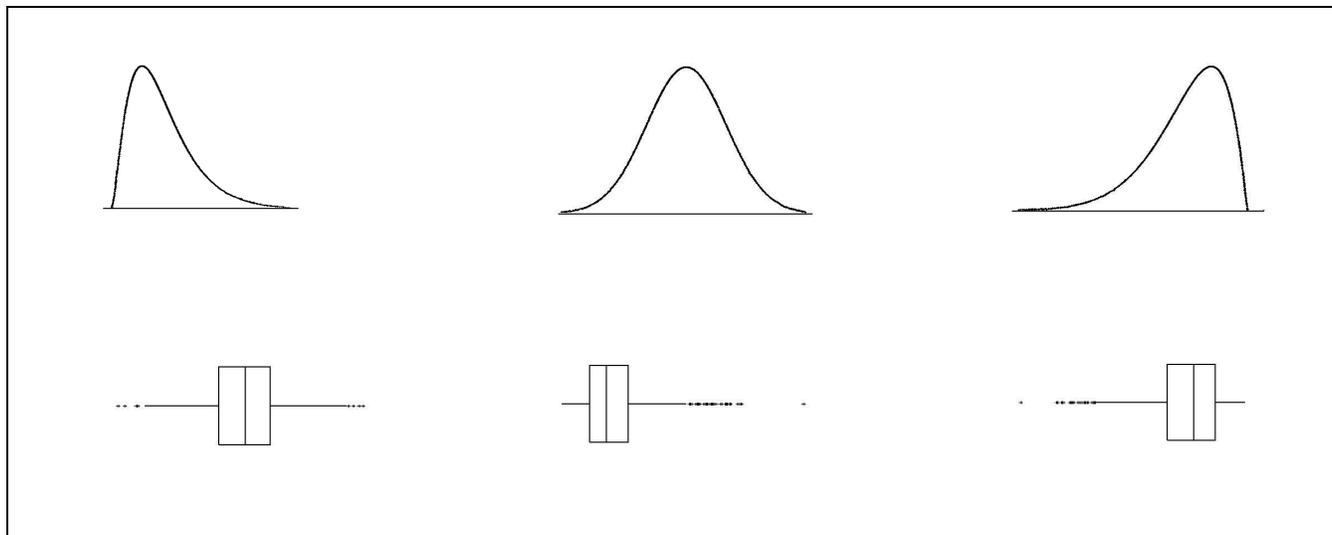
Abaixo de 11 s.m. estão todas as observações; portanto, o último ponto do gráfico corresponde ao par (11; 100).

Figura 6.1.3:



Exercício 2: Associe cada esboço de distribuição de freqüências ao seu box-plot correspondente na Figura 6.2.1 abaixo.

Figura 6.2.1:



Exercício 3: O gráfico box-plot fornece uma representação adequada da distribuição de freqüências de uma variável com distribuição bimodal ? Justifique.

Exercício 4: Suponha que receba propostas de emprego de duas empresas, A e B. Sabe-se que:

- A média dos salários na empresa A é R\$2000,00 e o terceiro quartil é R\$1000,00;
- A média dos salários na empresa B é R\$1000,00, o desvio padrão é R\$250,00 e a distribuição dos salários é simétrica em torno da média.

Suponha que, uma vez aceita a proposta de uma das empresas, seu salário será escolhido aleatoriamente entre todos os salários desta empresa. Você quer minimizar a probabilidade de ganhar menos que R\$1000,00, seu salário atual, ao mudar para uma destas duas empresas. Qual delas, empresa A ou B, você escolheria ? Explique.

Seção 7: Medidas de Posição II

Escores Padronizados

Exercício 1: Os 40 alunos da oitava série de uma escola pública foram submetidos, pelo seu professor de Educação Física, a quatro testes de aptidão física e a um teste de conhecimento desportivo. A média e o desvio padrão em cada teste para os 40 alunos e os resultados de três deles são mostrados no Quadro 7.1.1.

Quadro 7.1.1:

Teste	Média	Desvio Padrão	Escores Originais			Escores Padronizados		
			Pedro	João	Manuel	Pedro	João	Manuel
Número de abdominais em 2 minutos	30	6	32	40	25			
Salto em extensão (centímetros)	155	23	146	140	135			
Suspensão braços flexionados (segundos)	50	8	35	70	30			
Distância percorrida em 12 min (metros)	1829	274	2256	1726	1511			
Conhecimento desportivo (0 a 100 pts)	75	12	75	95	62			

- a) Complete a tabela com escores padronizados dos alunos Pedro e João em cada um cinco testes (Use apenas uma casa decimal).
- b) Para cada teste, indique o aluno com o pior desempenho e o aluno com o melhor desempenho. Que medida você precisa usar para responder ?

Quadro 7.1.2:

Teste	Aluno com o pior desempenho	Aluno com o melhor desempenho
Nº de abdominais em 2 min		
Salto em extensão (centímetros)		
Suspensão braços flexionados (segundos)		
Distância percorrida em 12 min (metros)		
Conhecimento desportivo (0 a 100 pts)		

- c) Para cada aluno, ordene os testes, do que ele teve seu pior desempenho até no que ele teve seu melhor desempenho. Que medida você deve usar responder ?

Quadro 7.1.3:

Aluno	Ordenação dos testes segundo desempenho do aluno				
	Pior	2º pior	...	2º melhor	Melhor
Pedro					
João					
Manuel					

- d) Suponha que o professor deseje ter um único número (índice), calculado para cada aluno, que expresse o desempenho global do aluno (em todos os testes). Que índice você sugere ? Calcule este índice para os alunos Pedro, João e Manuel.

Exercício 2: Um pesquisador, desejando classificar os bairros de uma cidade de acordo com a qualidade de vida destes bairros, criou o seguinte índice (Índice de Qualidade de Vida), calculado para cada bairro:

$$\begin{aligned}
 IQV = & \textit{Mediana da renda domiciliar (em reais) per capita} + \\
 & + \textit{proporção de casas com abastecimento de água e sistema de esgoto} + \\
 & + \textit{proporção de casas com energia elétrica} + \\
 & + \textit{proporção de ruas asfaltadas} + \\
 & + \textit{número de linhas de ônibus} + \\
 & + \textit{número de escolas} + \\
 & + \textit{número de hospitais e/ou postos de saúde}.
 \end{aligned}$$

- a) Do ponto de vista estatístico, por que está errado simplesmente somar todas estas variáveis ?
- b) Para criar o índice, o que o pesquisador deveria fazer com as variáveis antes de somá-las ?

Exercício 3: Considere o item (d) do Exercício 1. Suponha que, na definição da medida para o desempenho global de cada aluno, o professor escolha calcular, para cada um dos 40 alunos, a média dos escores padronizados em todos os testes. Entretanto, ele deseja dar mais peso àqueles testes com maior poder de discriminação entre os 40 alunos. Assim, ele deve utilizar uma média ponderada usando que valor como peso para cada teste ?

Exercício 4: Considerando novamente o Exercício 1, outro teste, natação, foi acrescentado aos cinco testes já mostrados. Neste teste, mediu-se o tempo (em segundos) gasto pelo aluno para nadar 100 metros. Note que, ao contrário dos testes anteriores, neste novo teste considera-se como melhor desempenho quanto menor for o valor do escore.

- a) Complete os quadros 7.4.1 a 7.4.3 referentes aos itens (a) a (c) do Exercício 1.

Quadro 7.4.1: Continuação item (a).

Teste	Média	Desvio Padrão	Escore Original			Escore Padronizado		
			Pedro	João	Manuel	Pedro	João	Manuel
Tempo gasto para nadar 100m(segundos)	30	5	22	30	45			

Quadro 7.4.2: Continuação item (b).

Teste	Aluno com o pior desempenho	Aluno com o melhor desempenho
Tempo gasto para nadar 100m (segundos)		

Quadro 7.4.3: Reformulação item (c).

Aluno	Ordenação dos testes segundo desempenho do aluno					
	Pior	2º pior	2º melhor	Melhor
Pedro						
João						
Manuel						

- b) Recalcule o índice de desempenho global de Pedro, João e Manuel que você sugeriu no item (d) do Exercício 1 considerando, também, o teste de natação.

Seção 8: Associação entre Variáveis Qualitativas

Exercício 1: Muitas pessoas acreditam que os réus que se declaram diante do júri culpados tendem a sofrer penas mais leves do que aqueles que se declaram inocentes. A Tabela 8.1.1 mostra os dados de uma amostra aleatória de acusados de roubo em São Francisco (EUA).

Tabela 8.1.1:

Sentença Final	Alegação Inicial do Acusado		Total
	Culpado	Inocente	
Enviado à prisão	392	58	450
Mantido em liberdade	564	14	578
Total	956	72	1028

- a) Complete as Tabelas 8.1.2 e 8.1.3 abaixo com as porcentagens calculadas por coluna e por linha, respectivamente. Interprete a porcentagem na primeira casela de cada tabela, observando a diferença de significado entre as duas.

Tabela 8.1.2:

Sentença Final	Alegação Inicial do Acusado		Total
	Culpado	Inocente	
Enviado à prisão			
Mantido em liberdade			
Total	100%	100%	100%

Tabela 8.1.3:

Sentença Final	Alegação Inicial do Acusado		Total
	Culpado	Inocente	
Enviado à prisão			100%
Mantido em liberdade			100%
Total			100%

- b) Identifique a variável explicativa e a variável resposta na hipótese do enunciado deste exercício.
- c) Considerando sua resposta ao item (b), qual das duas tabelas, 8.1.2 ou 8.1.3, deve ser utilizada para verificar tal hipótese? Justifique.
- d) Você diria que a sentença final é independente da alegação inicial do acusado nestes dados?

Exercício 2: A Tabela 8.2.1 abaixo apresenta a classificação de 200 pessoas condenadas à prisão por roubo ou furto, quanto ao sexo e à severidade da sentença recebida.

Tabela 8.2.1:

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	40	60	100
Feminino	60	40	100
Total	100	100	200

- a) Calcule as porcentagens de sentença “moderadas” e “severas” separadamente para homens e mulheres na Tabela 8.2.1.
- b) Analisando os resultados do item (a), você diria que existe associação entre o sexo do condenado e a severidade de sua sentença nestes dados? Explique.

As Tabelas 8.2.2 a 8.2.4 a seguir apresentam a classificação dos 200 condenados da Tabela 8.2.1 por sexo e severidade da sentença separadamente para cada uma das três categorias da variável *número de condenações anteriores*.

Tabela 8.2.2: Nenhuma condenação anterior.

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	10	0	10
Feminino	50	20	70
Total	60	20	80

Tabela 8.2.3: Uma ou duas condenações anteriores.

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	12	8	20
Feminino	10	10	20
Total	22	18	40

Tabela 8.2.4: Três ou mais condenações anteriores.

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	18	52	70
Feminino	0	10	10
Total	18	62	80

- c) Calcule as porcentagens de sentença “moderadas” e “severas” separadamente para homens e mulheres em cada uma das tabelas 8.2.2, 8.2.3 e 8.2.4.
- d) Analisando os resultados do item (c), você diria que existe associação entre o sexo do condenado e a severidade de sua sentença em cada estrato de número de condenações anteriores? O que mudou em sua conclusão do item (b)?

Seção 9: Associação entre Variáveis Quantitativas

Exercício 1: Associe a cada diagrama de dispersão abaixo o seu respectivo valor do coeficiente de correlação linear de Pearson dentre: a) 0,842 b) -0,787 c) 0,141

Figura 9.1.1: Diagrama de Dispersão 1:

$r =$ _____

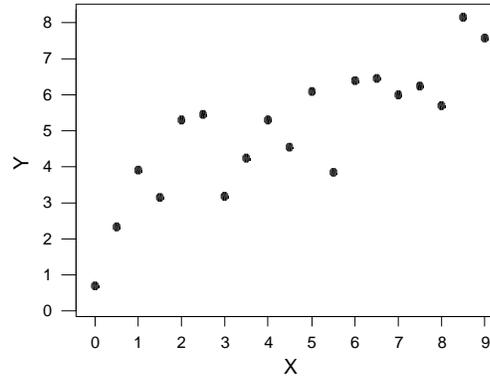


Figura 9.1.2: Diagrama de Dispersão 2:

$r =$ _____

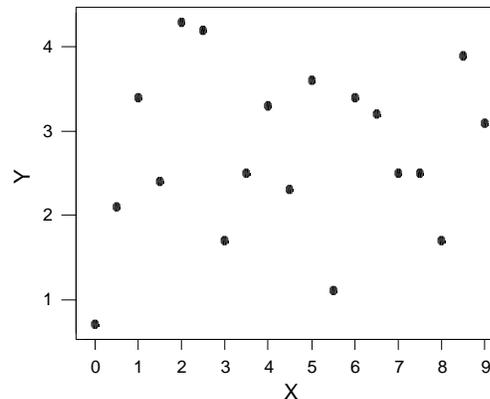
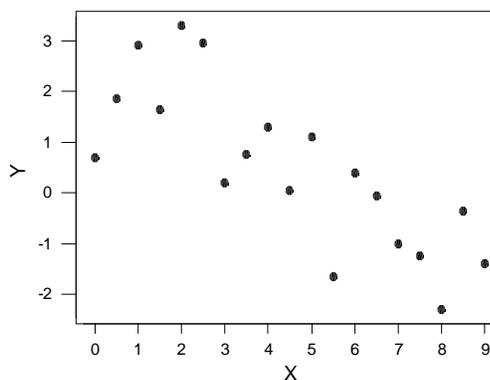


Figura 9.1.3: Diagrama de Dispersão 3:

$r =$ _____



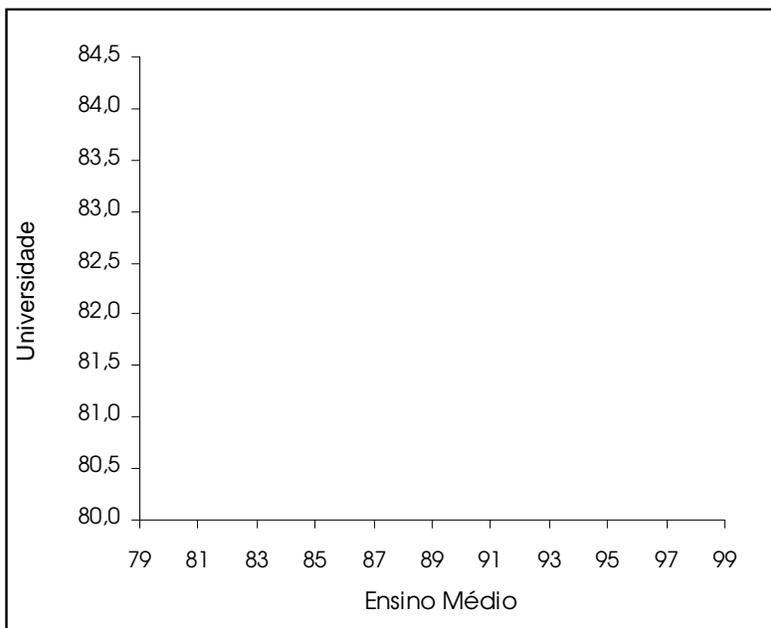
Exercício 2: O diretor de uma universidade está interessado em prever o desempenho do estudante na Universidade a partir de seu desempenho no Ensino Médio. Para saber se o desempenho do estudante no Ensino Médio é um bom preditor do seu desempenho na Universidade, ele selecionou aleatoriamente 15 alunos da Universidade e observou suas notas médias (escala de 0 a 100 pontos) no Ensino Médio e na Universidade, mostradas no Quadro 9.2.1, onde X representa o desempenho do estudante no Ensino Médio e Y o desempenho do estudante na Universidade.

Quadro 9.2.1:

I	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	80,0	81,0	6400	6561,0	6480,0
2	82,0	81,0	6724		
3	84,0	82,0		6724,0	
4	85,0	81,4			6919,0
5	87,0	82,1	7569	6740,4	7142,7
6	88,0	81,7	7744		
7	88,0	82,0		6724,0	
8	89,0	83,5			7431,5
9	90,0	83,1	8100	6905,6	7479,0
10	91,0	82,4	8281		
11	91,0	82,7		6839,3	
12	92,0	83,0			7636,0
13	94,0	83,9	8836	7039,2	7886,6
14	96,0	83,9	9216		
15	98,0	84,0		7056,0	
Soma		1237,7		102141,6	110220,9

- a) Identifique a variável explicativa e a variável resposta. Como você espera que seja a relação entre elas ?
- b) Desenhe o diagrama de dispersão na Figura 9.2.1 e comente sobre a relação entre o desempenho do aluno na Universidade e seu desempenho no Ensino Médio.

Figura 9.2.1: Diagrama de Dispersão do Desempenho na Universidade versus Desempenho no Ensino Médio.



c) Complete o Quadro 9.2.1, calcule e interprete o coeficiente de correlação linear de Pearson:

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2} \times \sqrt{n(\sum y_i^2) - (\sum y_i)^2}}$$

d) Calcule os coeficientes a e b da equação de regressão linear simples:

$$\text{Desempenho Universidade} = a + b \times \text{Desempenho Ensino Médio}$$

usando as equações

$$b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} \quad \text{e} \quad a = \bar{y} - b\bar{x}$$

e) Escreva a equação da reta, interprete seus coeficientes e desenhe-na na Figura 9.2.1.

f) Qual seria a estimativa, de acordo com a equação de regressão, para o desempenho na Universidade de um aluno que tivesse desempenho igual a 90 no Ensino Médio ? E para um aluno que tivesse desempenho igual a 65 no Ensino Médio ?

Exercício 3: Os biólogos estudam ursos anestesiando-os para obter medidas vitais como peso, altura e idade. Como os ursos, em sua maioria, são bastantes pesados e difíceis de serem levantados no meio da selva, será que os biólogos poderiam determinar o peso de um urso a partir de outras medidas mais fáceis de serem tomadas, como o perímetro de seu tórax, obtido apenas com o uso de uma fita métrica? Considere o experimento feito com oito ursos machos. Depois que cada urso foi anestesiado, os biólogos mediram o perímetro de seu tórax e obtiveram seu peso através de uma balança. Os resultados das medidas são mostrados no Quadro 9.3.1, sendo:

X = perímetro do tórax (em polegadas) e Y = peso (libras).

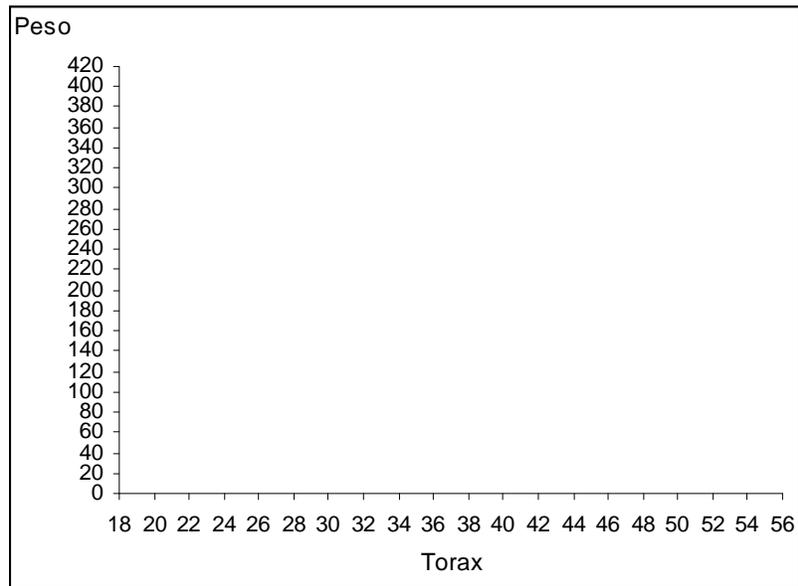
Quadro 9.3.1:

I	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	26	90	676	8100	2340
2	45	344	2025	118336	15480
3	54	416	2916	173056	22464
4	49	348	2401	121104	17052
5	41	262	1681	68644	10742
6	49	360	2401	129600	17640
7	44	332	1936	110224	14608
8	19	34	361	1156	646
Σ	327	2186	14397	730220	100972

a) Identifique a variável explicativa e a variável resposta. Como você espera que seja a relação entre elas ?

b) Desenhe o diagrama de dispersão na figura abaixo e comente sobre a relação entre o perímetro do tórax e o peso do urso.

Figura 9.3.1: Diagrama de Dispersão do perímetro do tórax versus o peso do urso.



c) Complete o Quadro 9.3.1, calcule e interprete o coeficiente de correlação linear de Pearson:

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2} \times \sqrt{n(\sum y_i^2) - (\sum y_i)^2}}$$

d) Calcule os coeficientes a e b da equação de regressão linear simples: **Peso = a + b × Tórax**

usando as equações

$$b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} \quad \text{e} \quad a = \bar{y} - b\bar{x}$$

e) Escreva a equação da reta, interprete seus coeficientes e desenhe-na na Figura 9.3.1.

f) Qual seria a estimativa para o peso de um urso que tivesse medida do tórax igual 30 polegadas ?
E para um urso com medida do tórax igual a 70 polegadas ?

Seção 10: Introdução à Amostragem

Exercício 1: Classifique o método de amostragem utilizado em cada uma das pesquisas descritas a seguir em:

- Amostragem Aleatória Simples (AAS);
- Amostragem Sistemática (AS);
- Amostragem Estratificada (AE), identificando a variável de estratificação;
- Amostragem por Conglomerados (AG), identificando os conglomerados.

Em todos os casos, identifique a população amostrada e a unidade de observação.

- a) Um jornalista escreve, em cartões separados, o nome de cada um dos 81 atuais senadores brasileiros, mistura-os e extrai 10 deles para uma entrevista sobre o problema da violação do painel eletrônico de votação.
- b) Um repórter da revista *EXAME* obtém uma relação numerada das 1000 empresas brasileiras que tiveram maiores cotações de ações na *BOVESPA* no ano de 2000, utiliza um computador para gerar 20 números aleatórios e então entrevista os gerentes gerais das empresas correspondentes aos números gerados.
- c) A empresa *Sony* seleciona cada 200º CD de sua linha de produção entre as 14 e 16 horas de um dia para fazer um teste de qualidade.
- d) Um entrevistador entrevista cada 10º eleitor que deixa uma seção eleitoral de uma cidade entre as 8 e 17 horas de um dia de eleições.
- e) O gerente geral de uma agência de um banco seleciona, por sorteio com igual chance, 50 correntistas com aplicações de até R\$5.000,00, outros 50 correntistas com aplicações entre R\$5.000,01 e R\$20.000,00 e outros 50 correntistas com aplicações entre R\$20.000,01 e R\$50.000,00 em 31 de dezembro de 2000 para receberem uma pesquisa via correio.
- f) Um médico entrevista todos os pacientes de leucemia internados em 20 hospitais selecionados por sorteio com igual chance dentre todos os hospitais de um estado.

Exercício 2: Comente a interpretação de “chance” ou “probabilidade” da personagem do seguinte texto:

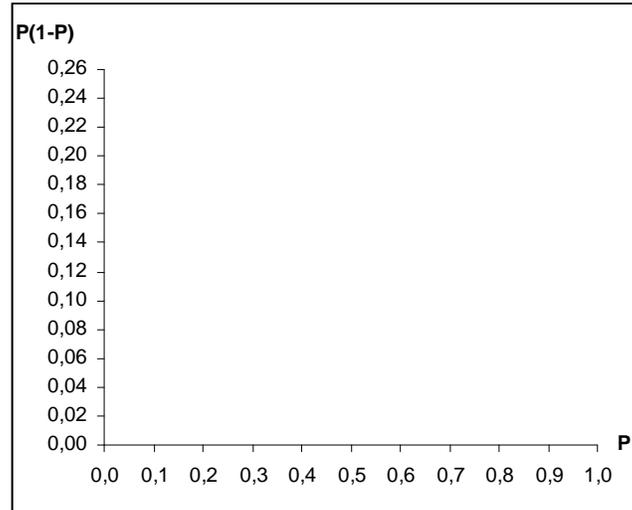
"Para determinar a causa de uma doença, uma pessoa teve que se submeter a um exame. Tudo estava indo bem, quando ela recebeu um formulário para assinar, no qual constava a seguinte advertência ao exame: 'Uma em cada 10.000 pessoas apresenta uma forte reação alérgica aos componentes químicos ingeridos durante este exame'. Muito preocupada, ela perguntou ao médico: - 'Em que número eles estão agora?'"

Exercício 3: Considere a fórmula do erro padrão da estimativa de uma proporção P de uma variável qualitativa nominal binária (ex: sexo masculino/feminino, aprova/rejeita lei, vota/não vota em certo candidato, etc.) em uma amostra aleatória simples de tamanho n retirada de uma população de tamanho N , para $n/N \approx 0$ (populações muito grandes):

$$s = \sqrt{\frac{P(1-P)}{n}}$$

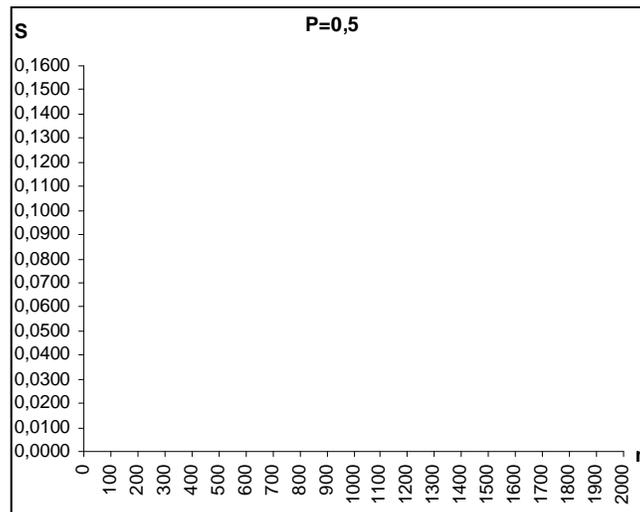
- a) Sabe-se que, na fórmula do erro padrão S , o numerador $P(1-P)$ quantifica a heterogeneidade da variável: quanto maior o valor $P(1-P)$, mais heterogênea é a variável na população. Complete a Figura 10.3.1 e comente sobre a relação entre a heterogeneidade da variável binária em função do valor da probabilidade P .

Figura 10.3.1:



- b) Escreva a equação do erro padrão S para a situação onde $P=0,5$, completando a Figura 10.3.2, que mostra a relação entre S e o tamanho da amostra n . Comente.

Figura 10.3.2:

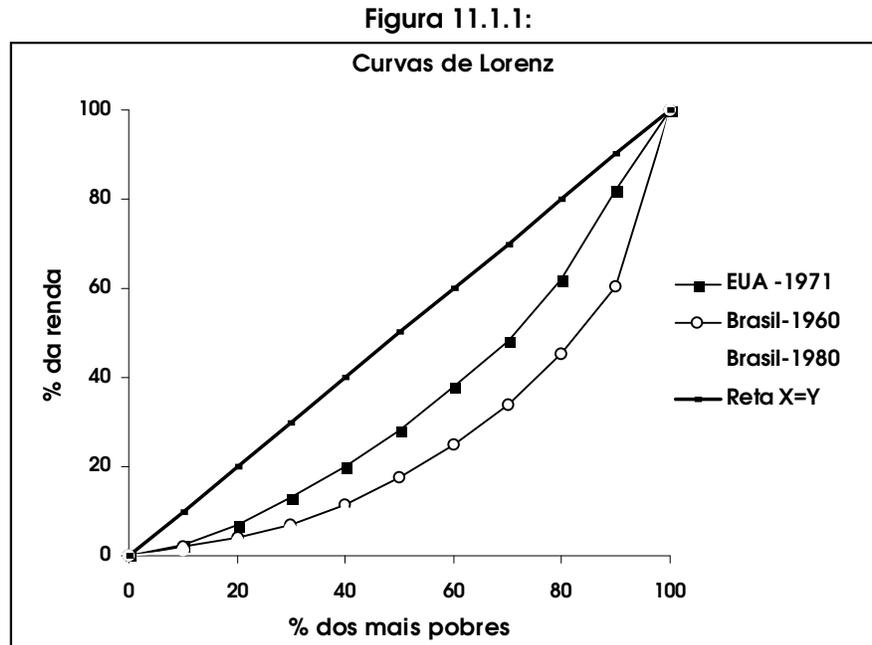


Seção 11: Medidas de Desigualdade

Exercício 1: A Figura 11.1.1 mostra a curva de Lorenz para a distribuição de renda da população economicamente ativa americana em 1971 e brasileira em 1960. Os dados de 1960 e 1980 da população brasileira estão no Quadro 11.1.1 a seguir.

Quadro 11.1.1: Distribuição de renda da população economicamente ativa brasileira (%)

% dos mais pobres	10	20	30	40	50	60	70	80	90	100
1960	1,9	3,9	6,9	11,3	17,4	24,9	33,9	45,2	60,4	100
1980	1,2	3,2	6,2	9,8	14,2	19,8	27,0	36,9	52,3	100



- Esboce a curva de Lorenz dos dados brasileiros de 1980 na Figura 11.1.1.
- O valor do índice de Gini para os dados de renda da população economicamente ativa no Brasil em 1980 é, aproximadamente, 0,10; 0,20; 0,60; 0,85 ou 0,90 ?
- Compare as três populações (EUA-1971, Brasil-1960 e Brasil-1980) quanto à distribuição de renda da população economicamente ativa.

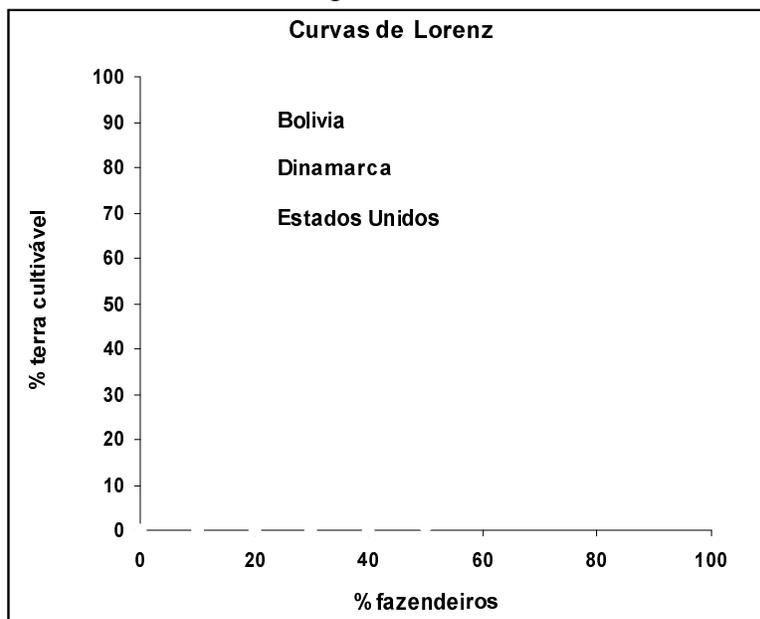
Exercício 2: A Tabela 11.2.1 mostra como a terra cultivável é distribuída entre fazendeiros na Bolívia, Dinamarca e nos Estados Unidos (Baseado em dados de B. Russett, World Handbook Social and Political Indicators, 1964).

- Desenhe a curva de Lorenz para cada um dos três países na Figura 11.2.1.
- Em que país a terra está mais igualmente distribuída entre os fazendeiros? Em que país a distribuição está mais desigual?

Tabela 11.2.1:

Porcentagem de fazendeiros	Porcentagem de terra cultivável		
	Bolivia	Dinamarca	Estados Unidos
0	0	0	0
10	0	6,0	2,5
20	0	12,0	5,0
30	0	18,0	7,5
40	0	24,0	10,0
50	1,0	30,0	13,0
60	1,6	36,0	18,0
70	2,2	45,0	22,0
80	3,0	54,0	28,0
90	4,0	70,0	42,0
100	100	100	100

Figura 11.2.1:



Exercício 3: A Tabela 11.3.1 mostra como a população foi dividida entre os 25 maiores países em 1963 e em 1968 (Dados da Comparative Analysis of Political Environments, como citada por Harf. Ledford, e Thomson, cit op.)

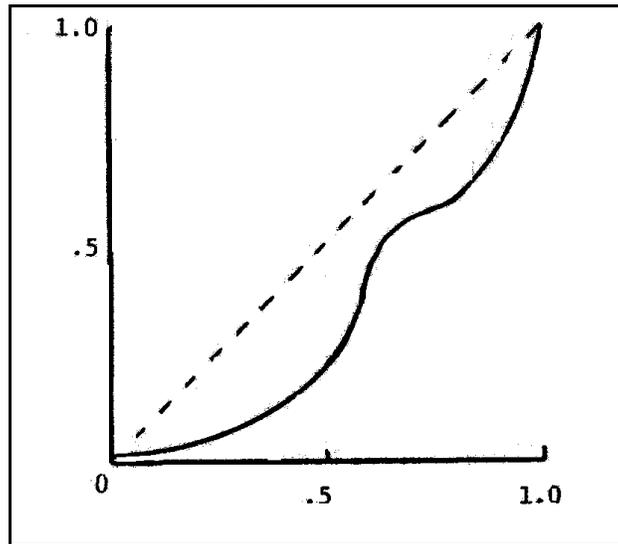
Tabela 11.3.1:

Porcentagem de Países	Porcentagem da População	
	1963	1968
0	0	0
20	4,6	4,8
40	10,5	10,7
60	19,3	19,4
80	34,5	34,7
100	100	100

- a) Houve muita mudança na distribuição da população entre 1963 e 1968?
- b) Como as curvas de Lorenz comparariam 1963 e 1968?
- c) Como você relaciona os dados na Tabela 11.3.1 (e o gráfico resultante) com o fato que a população do mundo está crescendo rapidamente?

Exercício 4: Uma curva de Lorenz pode ter a forma mostrada na Figura 11.4.1? Explique.

Figura 11.3.1:



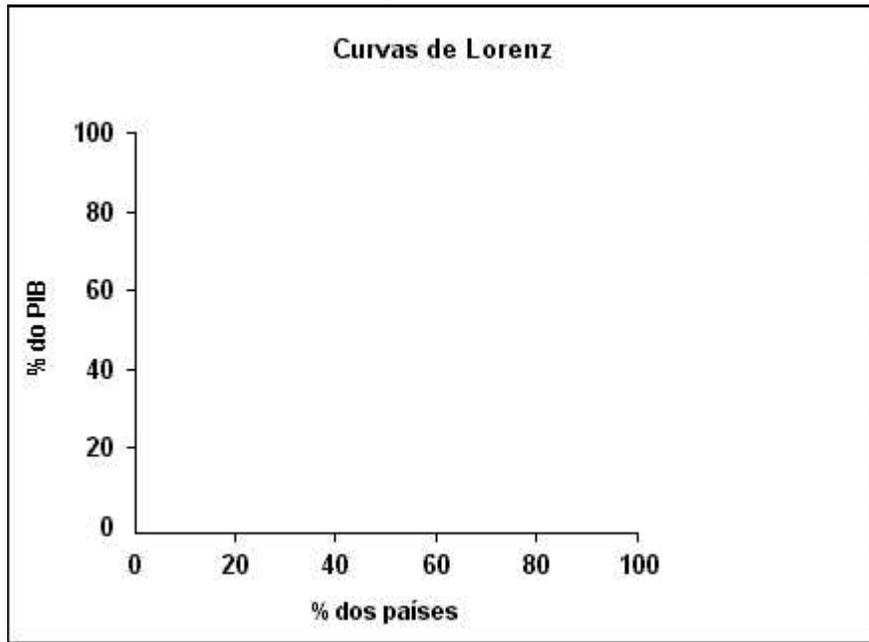
Exercício 5: A Tabela 11.5.1 mostra como produto interno bruto (PIB) é distribuído entre as nações do mundo.

Tabela 11.5.1:

% dos Países	% do PIB
0	0
10	0,1
20	0,2
30	0,5
40	1,0
50	1,8
60	2,8
70	5,8
80	11,0
90	21,0
100	100

- a) Desenhe a curva de Lorenz para os dados da Tabela 11.5.1 na Figura 11.5.1.
- b) Dezenhe na mesma figura a curva de Lorenz que corresponderia a uma distribuição *completamente desigual* do PIB e a curva de Lorenz que corresponderia a uma distribuição *completamente igual* do PIB.
- c) Com base na figura, você diria que o PIB é distribuído igualmente ou desigualmente entre os países do mundo?

Figura 11.5.1:



Solução da Seção 1: Abusos da Estatística

Exercício 1:

- a) Renda original: $R_0 = 10.000$
Corte de 10% sobre renda original: corte = $0,10 \times 10.000 = 1.000$
Renda após o corte de 10%: $R_1 = R_0 - \text{corte} = 10.000 - 1.000 = 9.000$
- b) O aumento de 10% será calculado sobre o salário atual de 9000 reais:
Aumento de 10%: $\text{aumento} = 0,10 \times R_1 = 0,10 \times 9.000 = 900$
Renda após aumento de 10%: $R_f = R_1 + \text{aumento} = 9.000 + 900 = 9.900$ (Renda Final)
- c) Desse modo, o corte de 10% seguido do aumento de 10% não restitui à funcionária o salário original de 10.000 reais. Ela tem uma perda de $10.000 - 9.900 = 100$ reais. Isto acontece porque o cálculo do aumento é feito sob o salário de 9000 reais após o corte e não sob o salário original de 10.000 reais.
- d) Para que o aumento posterior reponha o salário original, ele deve ser de $100/900 \approx 11,11\%$.

Exercício 2:

- a) Removendo-se 100% de uma quantidade resta zero (ou nada). Se o tal anti-séptico bucal reduz as placas dos dentes em 100%, significa que não vai restar área da arcada dentária com placa.
- b) Não tem significado algum. O máximo de redução que se pode ter é de 100%.

Observação: "Desconto" ou "redução" podem ser de no máximo 100%. Mas "aumento" ou "acrécimo" podem ser de qualquer porcentagem maior que zero.

Exemplo: Uma mercadoria custa originalmente R\$200.

- Desconto de:
 - $10\% = R\$200 \times 0,10 = R\20 Preço com desconto = $R\$200 - R\$20 = R\$180$
 - $50\% = R\$200 \times 0,50 = R\100 Preço com desconto = $R\$200 - R\$100 = R\$100$ (metade do preço)
 - $100\% = R\$200 \times 1,00 = R\200 Preço com desconto = $R\$200 - R\$200 = R\$0$ (de graça)
- Aumento de:
 - $10\% = R\$200 \times 0,10 = R\20 Preço com aumento = $R\$200 + R\$20 = R\$220$
 - $50\% = R\$200 \times 0,50 = R\100 Preço com aumento = $R\$200 + R\$100 = R\$300$
 - $100\% = R\$200 \times 1,00 = R\200 Preço com aumento = $R\$200 + R\$200 = R\$400$ (dobrou)
 - $200\% = R\$200 \times 2,00 = R\400 Preço com aumento = $R\$200 + R\$400 = R\$600$ (triplicou)

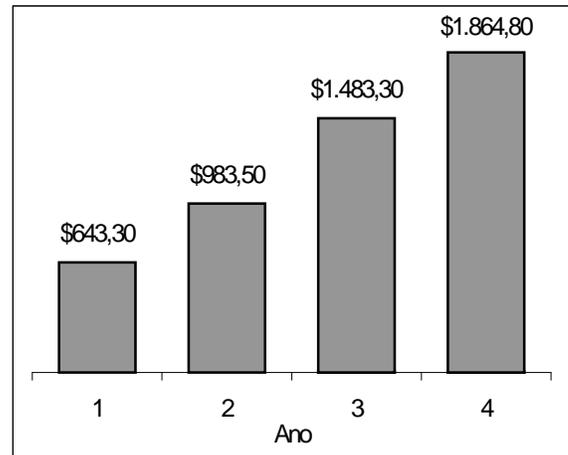
Exercício 3:

As barras não partem do valor zero no eixo vertical, exagerando, assim, a diferença no aumento das despesas entre os anos consecutivos, especialmente entre os anos 1 e 4.

Por exemplo, a impressão de que o aumento de despesas do ano 1 para o ano 4 foi de mais de 5 vezes (diferença visual), mas na verdade foi de $1864,8/643,5 \approx 3$ vezes.

O gráfico correto, partindo do zero no eixo vertical, é mostrado ao lado.

Figura 1.3.2: Gráfico correto.



Exercício 4:

- 1) A manchete faz supor que a pesquisa foi feita entre os moradores da cidade, quando na verdade ela foi feita apenas com os leitores do jornal;
- 2) A amostra foi auto-selecionada e assim, pode haver vício de seleção (leitores que respondem geralmente são aqueles mais interessados no assunto, então são aqueles que vão participar da amostra);
- 3) A amostra tem tamanho pequeno: $n = 20$;
- 4) Com $20 = 100\%$, apenas porcentagens múltiplas de 5 são possíveis: 0% (nenhum leitor), 5% (1 leitor), 10% (2 leitores), ..., 100% (todos os leitores);
- 5) A pergunta foi formulada de forma tendenciosa, induzindo as pessoas a responderem "sim", pois usa palavras fortes e apelativas com "matar", "milhões" e "inocentes".

Exercício 5:

Em geral, pessoas mais velhas (com 50 anos ou mais) tendem a arredondar a sua idade para um número perto de 0 ou 5, por exemplo: 50, 55, 60, 65 etc. Outras vezes este arredondamento é feito por outro membro da família que não sabe ao certo a idade das pessoas da casa com mais de 50 anos de idade. Isso dificilmente vai acontecer com idades abaixo de 50 anos, especialmente crianças e adolescentes.

Exercício 6:

- a) A primeira versão da pergunta é muito confusa, pois apresenta uma dupla negativa nos termos "impossível ... nunca", resultando em uma inversão dos significados das respostas positiva e negativa. Compare:

Resposta positiva: Acho possível que a exterminação de judeus pelos nazistas nunca tenha existido.
Com significado negativo: o holocausto não ocorreu.

Resposta negativa: Acho impossível que a exterminação de judeus pelos nazistas nunca tenha existido. Com significado positivo: o holocausto ocorreu.

Em uma pergunta oral, esta inversão pode não ser percebida pelo respondente. Além disso, observe que, apesar da pergunta englobar as duas alternativas de resposta (*possível e impossível*) em seu texto, foi usada uma palavra que indica certeza (*impossível*) e outra que indica a

possibilidade com incerteza (*possível*). Ou seja, as duas alternativas de respostas não aparecem com igual peso na pergunta.

b) A segunda versão parece menos confusa por eliminar a dupla negativa, mas cria outra distorção: apresenta não apenas uma, mas duas perguntas distintas:

1ª - *Acha possível que a exterminação de judeus pelos nazistas jamais ocorreu ?*

2ª - *Está certo de que (a exterminação de judeus pelos nazistas) realmente aconteceu ?*

Se o entrevistado responder "sim", não se saberá se ele respondeu:

- *Sim, eu acho possível que a exterminação de judeus pelos nazistas jamais ocorreu.*

ou se ele respondeu:

- *Sim, eu estou certo de que a exterminação de judeus pelos nazistas realmente aconteceu.*

A mesma confusão ocorrerá se o entrevistado responder "não".

Além disso, é a elaboração é tendenciosa, pois na frase que apresenta a alternativa de que a exterminação não ocorreu é usada uma palavra de incerteza ("possível") e para a frase que apresenta a alternativa de que a exterminação ocorreu é usada uma palavra de certeza ("certo"). Aqui também vale a observação colocada no item (a) sobre o uso de palavras com significado de incerteza para definir posições sobre dado assunto.

c) Uma forma mais clara de se fazer esta pergunta pode ser:

"Para você, a exterminação de judeus pelos nazistas ocorreu ou não ocorreu ?".

Note que, neste caso, não há inversão dos significados das respostas positiva e negativa.

Exercício 7:

Esta situação mostra um dos abusos mais comuns em pesquisa de opinião: perguntas tendenciosas, ou seja, formuladas com o propósito de induzir o entrevistado a responder a alternativa mais apropriada ao pesquisador. Neste exemplo, o pesquisador deseja induzir os entrevistados a responder que "sim, as pesquisas de opinião são válidas" usando o apelo dramático do desemprego (se ele disser que não acha que estas pesquisas sejam válidas, o entrevistador perde o emprego).

Além disso, foram feitas duas perguntas ao invés de uma: "As pesquisas de opinião são válidas" e "Vocês querem que eu perca meu emprego?". Se o entrevistado responder "sim" (ou "não"), não saberíamos para qual pergunta ele respondeu com esta afirmativa (ou negativa).

Exercício 8:

A definição correta da eficiência de um cursinho em termos de aprovação não deve ser feita através da porcentagem dos aprovados que pertence a tal cursinho e sim da porcentagem de alunos daquele cursinho que foram aprovados:

Cursinho	Aprovados	% de Aprovação
Alpha	1600	$1600/4000 = 0,40$ ou 40%
Beta	400	$400/500 = 0,80$ ou 80%
Gama	2400	$2400/6000 = 0,40$ ou 40%
Nenhum	600	$600/39500 = 0,02$ ou 2%

Assim, o cursinho com maior porcentagem de aprovação foi o Beta, pois 80% de seus alunos foram aprovados. O cursinho Gama, embora tenha um número absoluto de alunos aprovados maior, teve uma eficiência de apenas 40%.

Note que o cursinho Gama tem o maior número de alunos dentre os três cursinhos. Assim, mesmo se todos os cursinhos tivessem a mesma eficiência em aprovação, ele teria o maior número absoluto de aprovados. Por exemplo, suponha que os três cursinhos fossem igualmente eficientes, com cerca de, digamos, 80% de aprovação. Assim, o número esperado de aprovados em cada cursinho seria:

Cursinho	Candidatos	Nº Esperado de Aprovados
Alpha	4000	$4000 \cdot 0,8 = 3200$
Beta	500	$500 \cdot 0,8 = 400$
Gama	6000	$6000 \cdot 0,8 = 4800$

Seguindo o raciocínio da propaganda, o cursinho Gama teria ficado com 64% das 5000 vagas. Note ainda que, com este total de 6000 alunos, o cursinho Gama poderia até mesmo ter apenas 1601 aprovados (27% de aprovação, a menor dentre os cursinhos) que ainda poderia usar o artifício da propaganda e dizer que é melhor porque garantiu mais vagas.

Exercício 9:

Como temos o mesmo número de pacientes em cada grupo de comparação – tamoxifeno e placebo, a saber, 6944 pacientes ($13388 \div 2$), podemos comparar diretamente o número de pacientes que desenvolveram câncer de mama em cada grupo: 85 no grupo das pacientes que tomaram o tamoxifeno e 154 no grupo das que tomaram o placebo.

Assim, comparado ao grupo que tomou a droga, o grupo que não a tomou teve $154 - 85 = 69$ casos a mais de câncer – o que significa $69/85 = 81\%$ a mais e não os 45% calculados pela revista. Ou seja, se a revista considera surpreendente um aumento de 45%, imagina se tivesse calculado certo e encontrado os reais 81%!

A conta feita pela revista para encontrar os tais 45% provavelmente foi $69/154 = 45\%$, ou seja, ela tomou como base o grupo final e não o inicial da comparação.

Deve-se observar que, se o número de pacientes em cada grupo – droga e placebo, não fosse o mesmo, não poderíamos comparar diretamente o número de casos de câncer em cada grupo e sim as proporções (ou porcentagens) de casos em cada grupo.

Exercício 10:

As duas representações estão corretas se considerarmos que a diferença visual corresponde à diferença real de que em 2000 o número de vacas leiteiras é três vezes maior que em 1950.

- Na representação I (Figura 1.10.1) a vaca de 2000 tem área três vezes maior que a vaca de 1950:

$$\left. \begin{array}{l} \text{Área da vaca em 2000} = \text{altura} \times \text{largura} = 1,6\text{cm} \times 2,1\text{cm} = 3,36 \text{ cm} \\ \text{Área da vaca em 1950} = \text{altura} \times \text{largura} = 0,9\text{cm} \times 1,2\text{cm} = 1,08 \text{ cm} \end{array} \right\} \frac{3,36}{1,08} = 3$$

- Na representação II (Figura 1.10.2) temos três vacas a mais em 2000 que em 1950, todas do mesmo tamanho.

Entretanto, é muito mais fácil e rápido perceber esta diferença de três vezes mais olhando para a representação II do que para a representação I. Além disso, estamos falando de uma contagem (número de vacas leiteiras) que é maior em um ano que no outro e, assim, melhor representar esta diferença mostrando que, para cada vaca leiteira em 1950, havia 3 em 2000.

Solução da Seção 2: Proporção, Porcentagem, Taxa e Razão

Exercício 1:

a)

Quadro 2.1.1: Distribuição do peso corporal de homem com 70 Kg.

Órgão	Peso (Kg)	Peso (%)
Músculos	30,0	42,9%
Esqueleto	10,0	14,3%
Sangue	5,0	7,1%
Gastrointestinal	2,0	2,9%
Fígado	1,7	2,4%
Cérebro	1,5	2,1%
Pulmões	1,0	1,4%
Outros	18,8	26,9%
Total	70,0	100%

b) O peso total do sangue é 5,0 kg. Então o peso das células do sangue é 40% de 5,0kg = $5,0 \times 0,40 = 2,0\text{Kg} = 2000$ gramas.

d) A porcentagem do peso total do corpo devida as células do sangue é

$$\frac{2,0\text{kg}}{70,0\text{kg}} \times 100 = 2,8\%.$$

e) O esqueleto representa 14,3% (= $10,0/70,0 = 1/7$) do peso total do corpo (PTC). Se o peso do esqueleto vale V quilos, então temos que

$$14,3\% \text{ de PTC} = V \Rightarrow 0,143 \times \text{PTC} = V \text{ ou seja } \frac{1}{7} \text{ PTC} = V \Rightarrow \text{PTC} = 7V$$

Exercício 2:

a) O setor de Serviços foi o único que apresentou crescimento (29%). Veja a taxa de decréscimo dos outros setores:

Mês/Ano	Indústria	Construção Civil	Comércio	Serviços	Serviços Domésticos	Outros
Março/1990	18	10	18	38	14	2
Março/2000	16	8	16	49	10	1
(De)crescimento	$\frac{16-18}{18} = -11\%$	$\frac{8-10}{10} = -20\%$	$\frac{16-18}{18} = -11\%$	$\frac{49-38}{38} = 29\%$	$\frac{10-14}{14} = -29\%$	$\frac{1-2}{2} = -50\%$

b) Calculando 49% de 1.453.000 temos que 711.970 pessoas estavam ocupadas no setor de Serviços em março de 2000: $1.453.000 \times 0,49 = 711.970$

c) Não é possível determinar qual seria a nova porcentagem da categoria Construção Civil porque não foi fornecido o número de desempregados em Março/2000.

d) O número total de ocupados em Março/1999 era $TO = 1.093.000$:
 $0,18 \times TO = 196.740 \Rightarrow TO = 196.740/0,18 = 1.093.000$

Exercício 3:

- a) Foram 45688 mulheres inscritas em um total de $45688 + 32014 = 77702$ inscritos, que corresponde a uma proporção de $45688/77702 \approx 0,58799$. Ou seja, foram cerca de
- 6 mulheres para cada grupo de 10 inscritos;
 - ou 59 mulheres para cada grupo de 100 inscritos (59%);
 - ou 588 mulheres para cada grupo de 1000 inscritos;
 - ou 5880 mulheres para cada grupo de 10000 inscritos.
- b) Razão de sexos = #mulheres/#homens = $45688/32014 \approx 1,427125633$. Ou seja, inscreveram-se cerca de:
- 14 mulheres para cada grupo de 10 homens;
 - ou 143 mulheres para cada grupo de 100 homens;
 - ou 1427 mulheres para cada grupo de 1000 homens;
 - ou 14271 mulheres para cada grupo de 10000 homens.

Exercício 4:

- a) Aplicando a fórmula *imposto = soma das alíquotas aplicadas à parcela do rendimento em cada nível*, temos:

- i) Rendimento = 875 reais \Rightarrow Imposto = (0% de 875) = 0 reais.
- ii) Rendimento = 1400 reais \Rightarrow Imposto = (0% de 900) + (15% de (1400-900)) = (15% de 500) = 75 reais.
- iii) Rendimento = 3000 reais \Rightarrow Imposto = (0% de 900) + (15% de 900) + (27,5% de (3000-1800))
 $= 0 + 135 + (27,5\% \text{ de } 1200) = 135 + 330 = 465$ reais.

- b) Aplicando a fórmula *imposto = alíquota sobre rendimento – desconto*, temos:

- i) Rendimento = 875 reais \Rightarrow Imposto = (0% de 875) = 0 reais.
- ii) Rendimento = 1400 reais \Rightarrow Imposto = (15% de 1400) – 135 = 210 – 135 = 75 reais.

que equivale ao cálculo anterior: $(0\% \text{ de } 900) + (15\% \text{ de } (1400-900)) =$
 $= (15\% \text{ de } 1400) - (15\% \text{ de } 900) = 210 - 135 = 75$ reais.

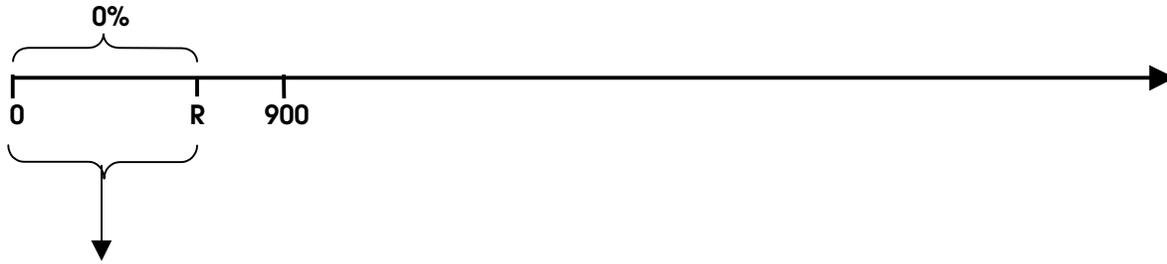
- iii) Rendimento = 3000 reais \Rightarrow Imposto = (27,5% de 3000) – 360 = 825 – 360 = 465 reais.

que equivale ao cálculo anterior: $(0\% \text{ de } 900) + (15\% \text{ de } 900) + (27,5\% \text{ de } (3000 - 1800)) =$
 $= 135 + (27,5\% \text{ de } 3000) - (27,5\% \text{ de } 1800)$
 $= 135 + 825 - 495 = 825 - 360 = 465$ reais.

Veja a explicação esquemática da equivalência entre as duas formas de cálculo na Figura 2.4.1 a seguir.

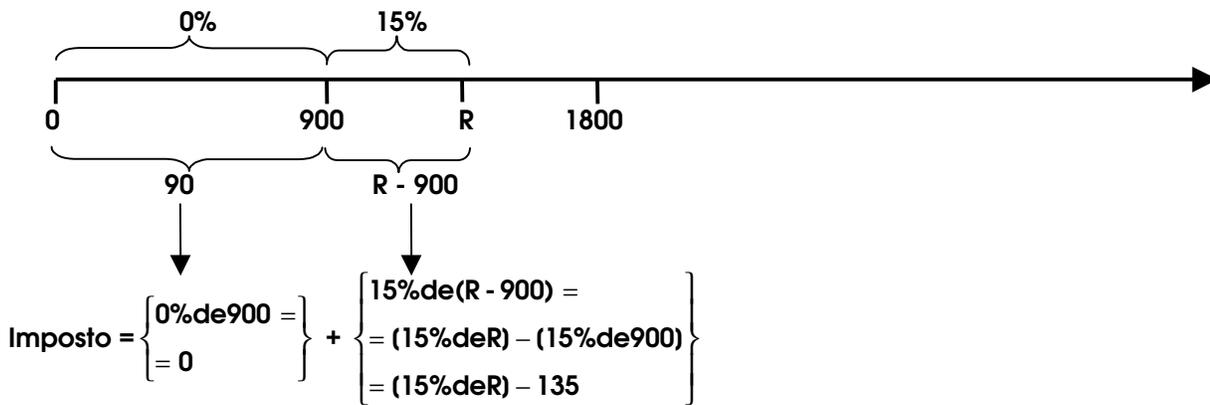
Figura 2.4.1: Esquema do cálculo do Imposto de Renda para Rendimentos de Pessoa Física.

a) Rendimento $R \leq 900$ reais:



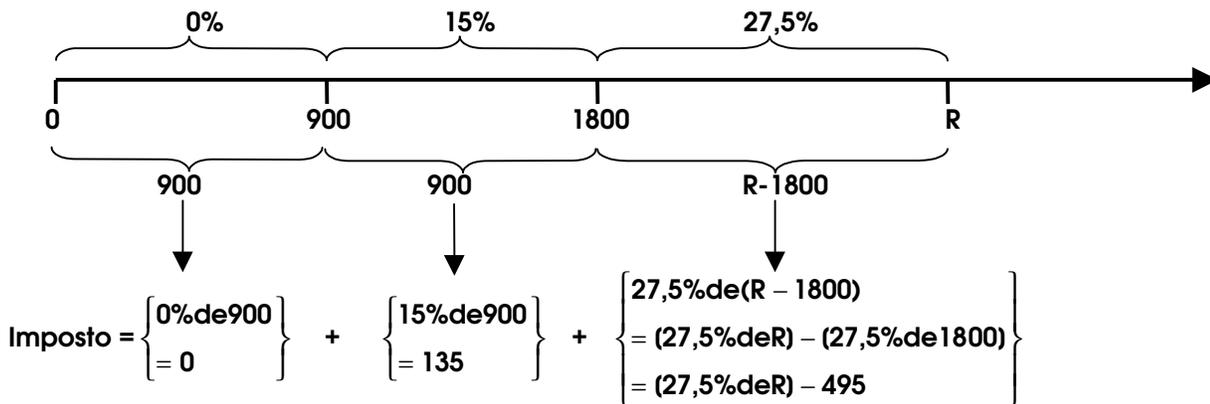
Imposto = 0% de $R = 0$ reais.

b) Rendimento $900 < R \leq 1800$ reais:



Imposto = $0 - 135 + (15\% \text{ de } R) = (15\% \text{ de } R) - 135$

c) Rendimento $R > 1800$ reais:



Imposto = $0 + 135 - 495 + (27,5\% \text{ de } R) = (27,5\% \text{ de } R) - 360$

Exercício 5:

i. Não se pode comparar diretamente o número absoluto de pedidos de isenção deferidos entre dois anos (ou ao longo dos seis anos), pois o número de pedidos recebidos em cada ano é diferente (aumentou) ao longo deste período.

Assim, deve-se comparar a *porcentagem de pedidos deferidos em tal ano em relação ao total de pedidos recebidos no mesmo ano* com a mesma porcentagem calculada para o outro ano de interesse. Por exemplo, em 1997, foram 4322 pedidos deferidos no total de 5285 recebidos, ou seja, $4322/5285 = 0,82 = 82\%$ de pedidos deferidos. Para os outros anos, temos:

Ano	1997	1998	1999	2000	2001	2002
Porcentagem de Pedidos de Isenção Deferidos	82%	72%	53%	51%	62%	58%

Comparando o primeiro ano como o último, como foi feito no comentário, houve, na verdade, uma redução do apoio da UFMG aos alunos carentes, pois, em 1997, 82% dos pedidos de isenção da taxa de inscrição foram deferidos, contra 58% no vestibular mais atual.

ii. Número de pedidos de isenção recebidos em 2001: 30425

Número de pedidos de isenção recebidos em 2002: 54234

$$\frac{54234 - 30425}{30425} = \frac{23809}{30425} = 0,78 \text{ ou } 78\% \text{ de aumento de 2001 para 2002.}$$

Número de pedidos de isenção deferidos em 2001: 18770

Número de pedidos de isenção deferidos em 2002: 31362

$$\frac{31362 - 18770}{18770} = \frac{12592}{18770} = 0,67 \text{ ou } 67\% \text{ de aumento de 2001 para 2002.}$$

iii. Nos dois últimos vestibulares, a UFMG não conseguiu acompanhar a demanda por isenção da taxa de inscrição: em 2001, a porcentagem de pedidos de isenção deferidos foi de 61,7%, enquanto em 2002 foi de 57,8%.

iv. A aprovação no Vestibular da UFMG é menor para os candidatos que conseguiram isenção da taxa de inscrição. No ano 2001, por exemplo, apenas 2,3% dos candidatos isentos foram aprovados, enquanto que, para os candidatos não isentos da taxa, a porcentagem de aprovação foi de 6,4%.

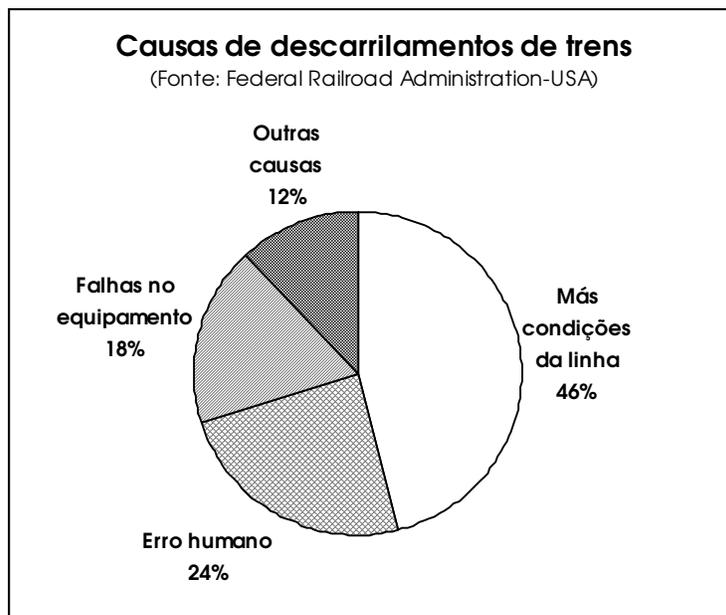
Solução da Seção 3: Tabelas e Gráficos

Exercício 1:

Tabela de Freqüências:

Causa do Descarrilamento	Freqüência Absoluta	Freqüência Relativa
Má condição da linha	23	46%
Erro humano	12	24%
Falhas no equipamento	9	18%
Outras causas	6	12%
Total	50	100%

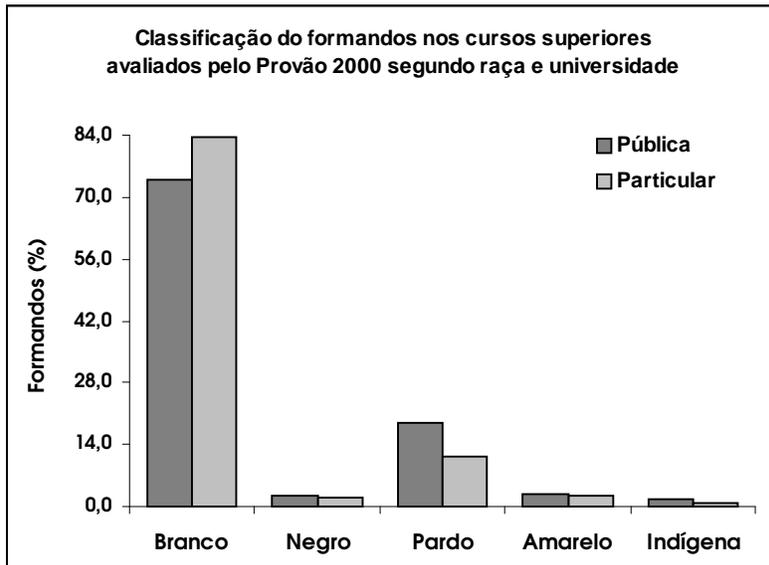
Gráfico em Setores:



Exercício 2:

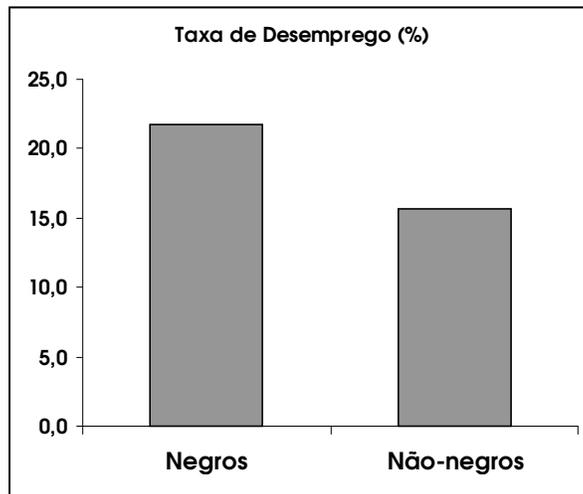
- A comparação entre as parcelas de cada raça nas universidades públicas e privadas através dos gráficos em setores (Figura 3.2.1) é prejudicada pelo fato de compararmos as áreas dos respectivos setores nos círculos, ou seja, nossos olhos têm que fazer a comparação de muitas quantidades bidimensionais simultaneamente. Além disso, a raça branca corresponde a um setor com grande área nos dois gráficos e, assim, as outras raças têm setores com áreas muito pequenas para serem comparadas visualmente.
- O gráfico de barras solicitado (Figura 3.2.2) é mostrado a seguir. Neste gráfico, é mais fácil visualizar as diferenças entre as parcelas de cada raça nas universidades públicas e privadas porque a comparação é feita através das alturas das barras, uma medida unidimensional. Além disso, as barras de cada raça estão lado a lado o que, adicionalmente, reduz o efeito da predominância da raça branca na visualização, pois se pode centrar a atenção apenas na raça que se está comparando a cada vez.

Figura 3.2.2: Gráfico em Barras Verticais.

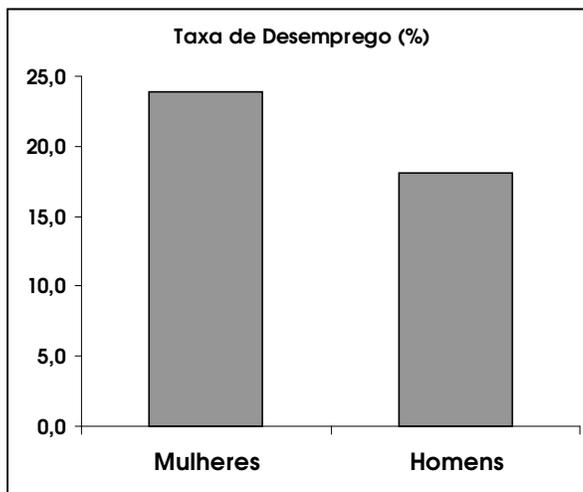


Exercício 3:

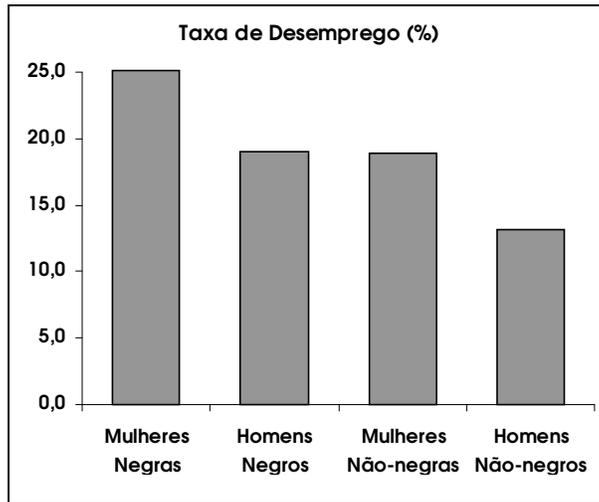
a) O efeito da raça na taxa de desemprego.



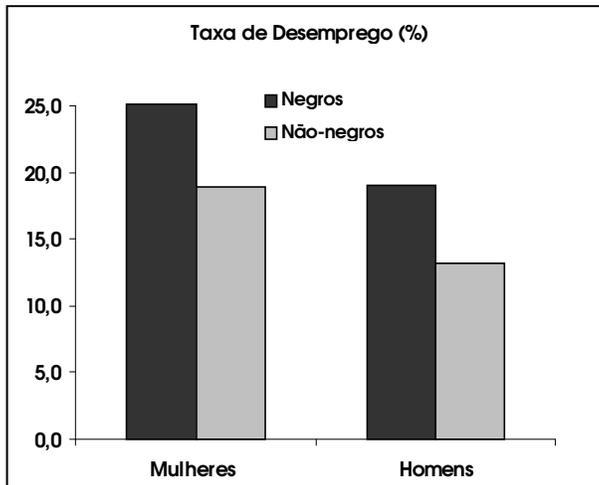
b) O efeito do sexo na taxa de desemprego.



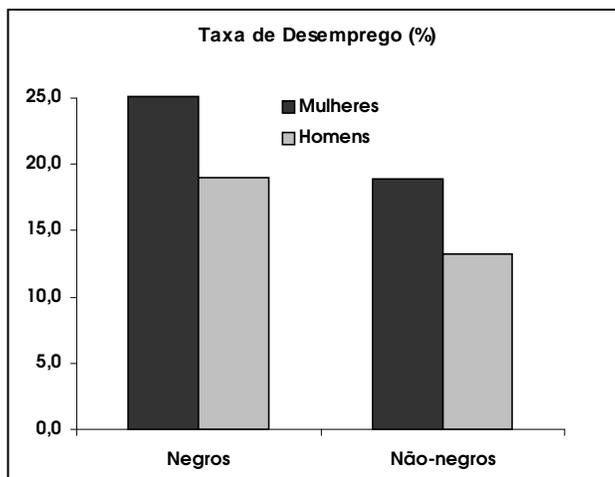
c) O efeito simultâneo da raça e do sexo na taxa de desemprego.



d) O efeito da raça na taxa de desemprego dentro de cada sexo.



e) O efeito sexo na taxa de desemprego dentro de cada raça.



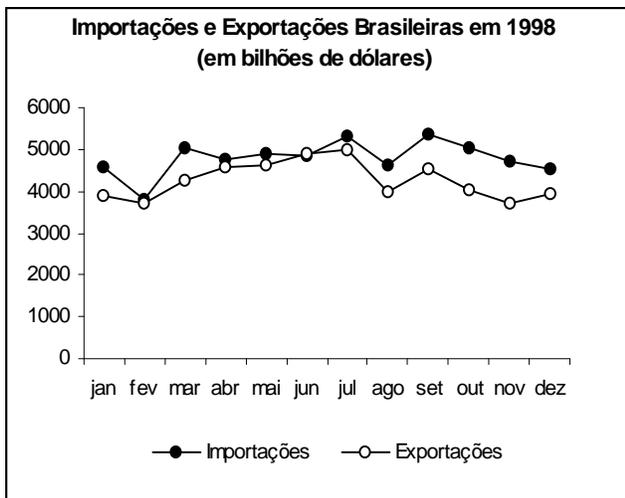
Exercício 4:

Como desejamos não somente descrever a evolução temporal das duas séries, como também comparar importações e exportações ao longo do tempo, faremos um gráfico de linhas conjunto para as duas séries temporais, importações e exportações, como está mostrado na representação gráfica I. Note que o eixo vertical (valor) começa no zero.

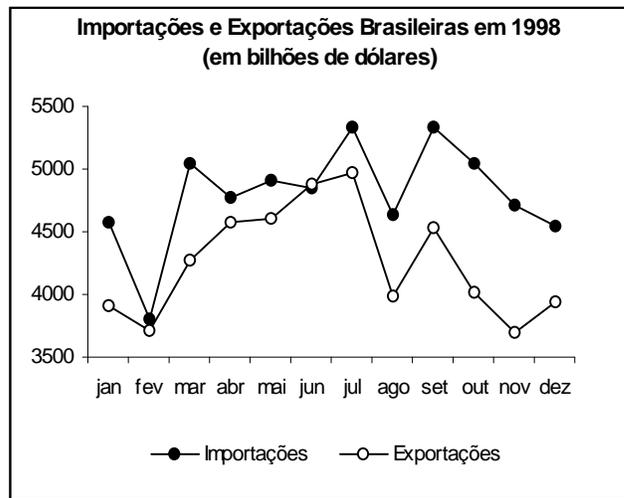
Podemos ver que tanto os valores das importações como das exportações permaneceram estáveis no período, ou seja, sem muita variação. As importações estiveram sempre acima das exportações, especialmente a partir de agosto/setembro.

A representação gráfica II mostra como ficaria a visualização das séries se o eixo vertical não começasse em zero. Teríamos impressão de grande variação em ambas as séries ao longo do período e um superdimensionamento das diferenças entre as duas séries.

Representação Gráfica I.



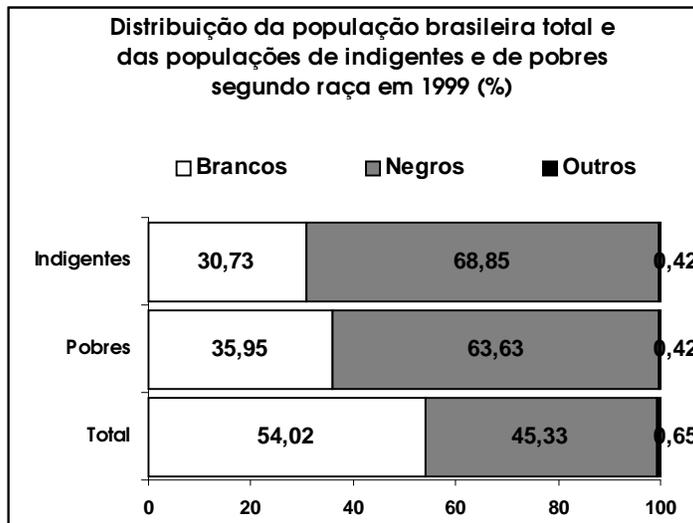
Representação Gráfica II.



Exercício 5:

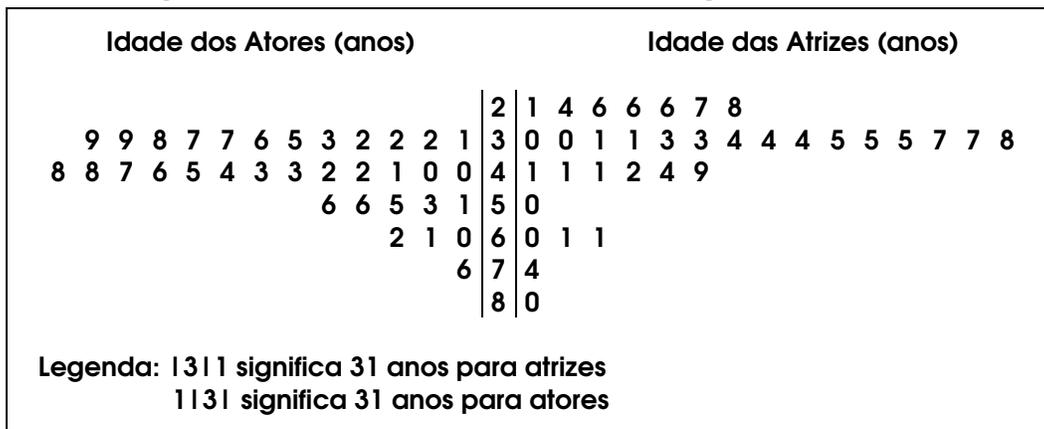
A composição racial (negros/brancos/outros) da população de pobres e indigentes pode ser representada em um gráfico de barras com as porcentagens de indivíduos de cada raça, como o mostrado a seguir. Como foi feito no texto, também é necessário informar no gráfico a composição racial de toda a população, pois sabemos que esta não está dividida igualmente entre as raças.

O gráfico mostra que a participação da população negra nas subpopulações de pobres e indigentes é muito maior que sua participação na população brasileira total, mostrando uma clara diferença racial na distribuição da renda.



Exercício 6:

Figura 3.6.1: Ramo-e-folhas para "Idade ao ganhar o Oscar".



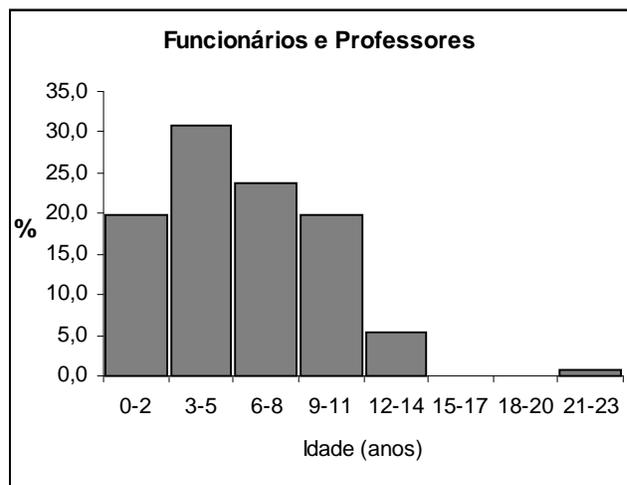
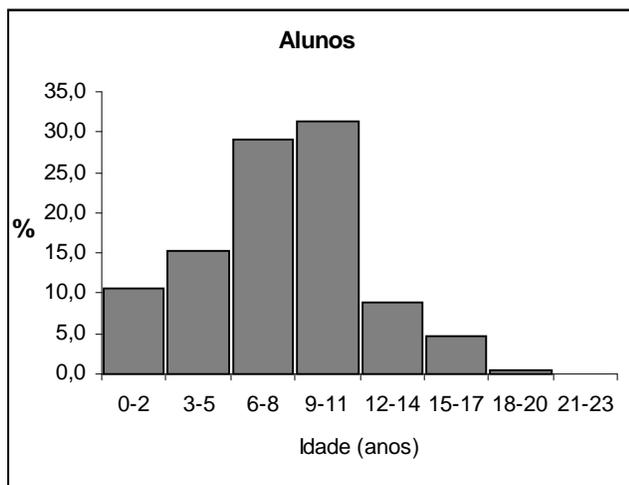
As atrizes ganham o Oscar mais cedo (na casa dos 30 anos) que os atores (na casa dos 40 anos). Há maior variabilidade entre atrizes do que entre atores na idade ao ganhar o Oscar. A distribuição desta variável é assimétrica à direita para ambos os sexos, com assimetria mais acentuada para o grupo das atrizes.

Exercício 7:

Tabela 3.7.1:

Idade (anos)	Estudantes		Funcionários e Professores	
	Freqüência Absoluta	Freqüência Relativa (%)	Freqüência Absoluta	Freqüência Relativa (%)
01-12	23	10,6	30	19,7
31-15	33	15,2	47	30,9
61-18	63	29,0	36	23,7
91-111	68	31,3	30	19,7
121-114	19	8,8	8	5,3
151-117	10	4,6	0	0,0
181-120	1	0,5	0	0,0
211-123	0	0,0	1	0,7
Total	217	100	152	100

Histogramas:



Os alunos têm, em geral, carros mais velhos (6 a 11 anos) que os funcionários e professores (3 a 8 anos). Há maior variabilidade entre os alunos do que entre os professores e funcionários quanto à variabilidade na idade de seus carros. A distribuição da idade dos carros é aproximadamente simétrica no grupo de alunos e assimétrica à direita no grupo de funcionários e professores. Note que há um carro de professor ou funcionário que tem idade muito maior que a maioria do grupo.

Exercício 8:

a)

Figura 3.8.1: Gráfico Circular.

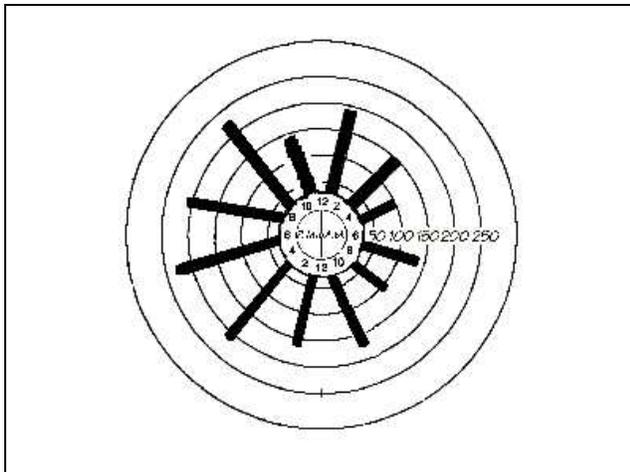
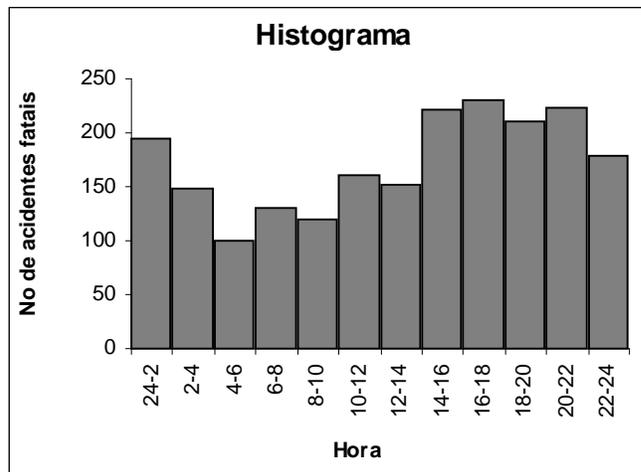


Figura 3.8.2: Histograma.



- b) O gráfico circular ilustra melhor os dados porque as classes, os horários neste caso, formam um ciclo. Note que, no histograma, há uma quebra deste ciclo. Temos a falsa impressão de que há dois picos de colisões: na primeira classe e em torno da nona classe. Entretanto, no gráfico circular vemos que na verdade há um só pico por volta de 16-18 horas, porque a primeira classe é também a 13ª classe.
- c) Não, porque não há dados sobre a quantidade de veículos motorizados que circulam nos diversos horários. Quando há mais veículos circulando, a exposição a acidentes é maior. Se não houvesse diferença na quantidade de veículos circulando entre os diversos horários, esperaríamos um número maior de colisões no horário com maior risco (perigo) de colisão. Com certeza no horário de 4 às 6 horas da manhã deve haver menos veículos circulando. Portanto, mesmo que o risco neste horário seja igual ao dos outros horários, era de se esperar um número menor de colisões.

Exercício 9:

- a) São sete as classes de renda familiar consideradas neste gráfico (Figura 3.9.1): US\$1 a US\$499; US\$500 a US\$999; US\$1000 a US\$1499; US\$1500 a US\$1999; US\$2000 a US\$2999; US\$3000 a US\$4999; US\$5000 ou mais.

As classes não têm a mesma amplitude; as quatro primeiras têm amplitude de US\$499, a quinta tem amplitude de US\$999, a sexta classe tem amplitude de US\$1999 e a última tem amplitude indefinida (classe aberta). Estas diferenças de amplitude intraclasse foram representadas no gráfico pela largura do espaço entre os limites de classe, proporcional à amplitude das classes.

- b) Cada símbolo representa 2% das famílias da cidade (por exemplo, se a cidade tem 100 mil famílias, cada símbolo representa duas mil famílias). Assim, para somar 100%, temos 50 símbolos no gráfico, dos quais 20,5 são coloridos de preto e 29,5 são coloridos de branco, ou seja, 41% das famílias são compostas por indivíduos negros e 59% por indivíduos brancos.
- c) A variável renda familiar tem distribuição acentuadamente assimétrica com grande concentração à direita, mostrando predominância de famílias com rendas baixas. Há grande variabilidade nas rendas: uma diferença de cinco vezes entre as menores e as maiores rendas.
- d) A distribuição da renda entre famílias negras e não-negras é desigual, pois as famílias negras se concentram nas menores faixas de renda, não havendo famílias negras com rendas altas.

e)

Tabela 3.9.1:

Renda (US\$)	Número de Símbolos*			Frequência Relativa (%)		
	Negros	Branços	Total	Negros	Branços	Total
1 1-1 499	16	4	20	32%	8%	40%
5001-1 999	4	6	10	8%	12%	20%
10001-1 1499	0,5	4,5	5	1%	9%	10%
15001-1 1999	0	5	5	0%	10%	10%
20001-1 2999	0	5	5	0%	10%	10%
30001-1 4999	0	4	4	0%	8%	8%
5000 ou mais	0	1	1	0%	2%	2%
Total	20,5	29,5	50	41%	59%	100%

* Convenção: um símbolo faz parte de uma classe quando sua base ("pés") faz parte desta classe.

- f) O gráfico de barras verticais (Figura 3.9.2) não representa as diferenças de amplitude entre as classes de renda, pois todas as barras foram desenhadas com a mesma largura. Isto resulta em um "achatamento" da distribuição de frequências da variável: ela aparenta menos acentuadamente assimétrica do que na realidade é. O gráfico de barras tem a vantagem de apresentar os valores das frequências relativas em cada classe, divididas entre negros e não-negros. Os dois gráficos representam igualmente bem a desigualdade de renda entre as raças, embora fique mais claro na Figura 3.9.1 que as frequências se relacionam a pessoas.

Exercício 10:

Tabela 3.10.1: Sexo X Severidade da Sentença.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino	40	60	100
Feminino	60	40	100
Total	100	100	200

Tabela 3.10.2: Sexo X Número de Condenações Anteriores.

Sexo	Condenações Anteriores			Total
	Nenhuma	Uma ou duas	Três ou mais	
Masculino	10	20	70	100
Feminino	70	20	10	100
Total	80	40	80	200

Tabela 3.10.3: Severidade da Sentença X Número de Condenações Anteriores.

Severidade Sentença	Condenações Anteriores			Total
	Nenhuma	Uma ou duas	Três ou mais	
Moderada	60	22	18	100
Severa	20	18	62	100
Total	80	40	80	200

Tabela 3.10.4: Sexo X Severidade da Sentença segundo Número de Condenações Anteriores.

a) Nenhuma condenação anterior.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino	10	0	10
Feminino	50	20	70
Total	60	20	80

b) Uma ou duas condenações anteriores.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino	12	8	20
Feminino	10	10	20
Total	22	18	40

c) Três ou mais condenações anteriores.

Sexo	Severidade da Sentença		Total
	Moderada	Severa	
Masculino	18	52	70
Feminino	0	10	10
Total	18	62	80

Solução da Seção 4: Medidas de Tendência Central

Exercício 1:

Tempo de espera na fila do banco	Média	Mediana	Moda
Banco A (fila única)	7,15	7,2	7,7
Banco B (fila múltipla)	7,15	7,2	7,7

A distribuição do tempo de espera tem o mesmo ponto de concentração (por volta de 7 minutos) tanto para fila única quanto para filas múltiplas. Note que a moda aqui não tem muita utilidade, pois, sendo a amostra muito pequena, dificilmente aparecerão muitas repetições do mesmo valor.

Exercício 2:

Largura máxima do crânio	Média	Mediana	Moda
4000 aC	128,7	128,5	131
150 dC	133,3	133,5	126

As larguras máximas dos crânios em 4000 aC são tipicamente menores do que em 150 aC. Note que a moda aqui não tem muita utilidade, pois, sendo a amostra muito pequena, dificilmente aparecerão muitas repetições do mesmo valor.

Exercício 3:

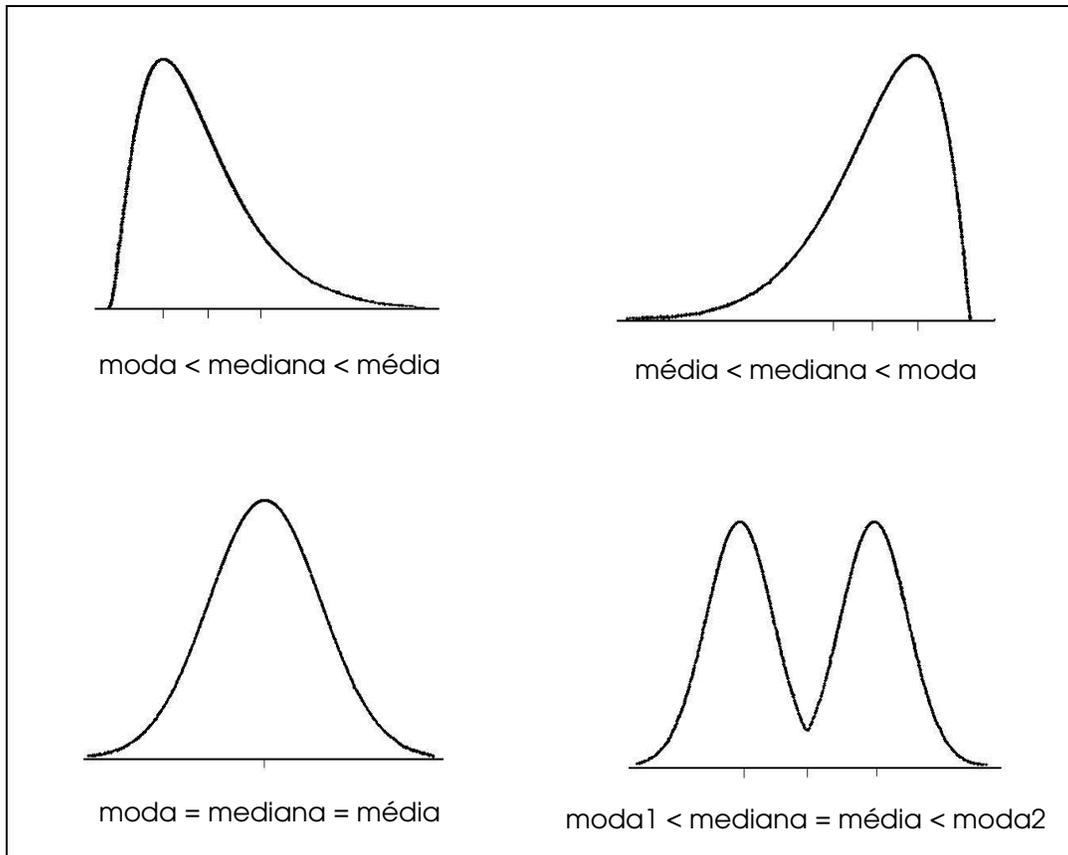
A mediana é a medida de tendência central mais adequada neste caso, pois ela pode ser calculada apenas com a informação de que estas três mulheres têm mais de 30 anos, já que precisamos somente da ordem dos dados para calcular a mediana e estas três informações faltantes estão no extremo da distribuição. O mesmo não acontece com a média. A média calculada com apenas os 12 valores informados poderia estar muito subestimada se estas três mulheres tiverem idades muito maiores que as observadas.

Exercício 4:

- a) Falsa. Correção: "mediana".
- b) Verdadeira.
- c) Falsa. Correção: "média".
- d) Falsa. Correção: "moda".
- e) Falsa. Correção: "assimétrica com cauda à esquerda".

Exercício 5:

Figura 4.5.1a:



Solução da Seção 5: Medidas de Variabilidade

Exercício 1:

Grupo 1: amplitude = $20 - 1 = 19$ desvio padrão = 5,7
Grupo 2: amplitude = $19 - 2 = 17$ desvio padrão = 6,7

- a) Os valores da amplitude sugerem que o Grupo 1 tem variabilidade maior.
- b) A amplitude é enganosa neste caso porque foi influenciada pelo valor discrepante do mínimo.
- c) O desvio padrão sugere que o Grupo 2 tem maior variabilidade e corresponde à realidade dos dados.

Exercício 2:

Quadro 5.2.2
Medidas de Tendência Central.

Tempo de espera	Média	Mediana	Moda
Banco A (fila única)	7,15	7,2	7,7
Banco B (fila múltipla)	7,15	7,2	7,7

Medidas de Variabilidade.

Tempo de espera	Desvio Padrão	Coefficiente de Variação	Amplitude Total
Banco A (fila única)	0,48	7%	1,2
Banco B (fila múltipla)	1,82	25%	5,8

Podemos ver que, embora o tempo médio (e mediano, e modal) de espera seja o mesmo nos dois tipos de fila (única e múltipla), a variabilidade é menor na fila única. Ou seja, é mais provável que se tenha de esperar um tempo muito acima da média na fila múltipla que na fila única.

Exercício 3:

Como as lâmpadas das duas marcas tenham o mesmo tempo médio de vida, escolheria as lâmpadas *Ultraligth* por terem menor desvio padrão. Isto porque embora as lâmpadas *Electrolyte* possam durar muito (p.ex. 3750h) elas também terão igual chance de durar muito pouco (p. ex. 2250h).

Note que você também pode usar o coeficiente de variação, ao invés do desvio padrão, para chegar a esta conclusão. Aqui, seu uso não é indispensável, pois se trata da mesma variável, medida na mesma escala e com médias iguais nos dois grupos.

Exercício 4:

Como as medidas dos dois aparelhos estão em escalas diferentes, devemos usar o coeficiente de variação para compará-los quanto à variabilidade.

$$\text{Aparelho 1: } CV_1 = \frac{0,019mm}{5,32mm} = 0,004 \text{ ou } 0,4\% \qquad \text{Aparelho 2: } CV_2 = \frac{0,03cm}{6,4cm} = 0,005 \text{ ou } 0,5\%$$

Portanto, o aparelho 1 é relativamente mais preciso, pois têm menor variabilidade em torno do valor correto.

Note que, neste caso, estamos supondo que a média amostral seja igual ou bem próximo ao verdadeiro valor da medida, que pode ser conhecido usando-se aparelhos mais precisos.

Exercício 5:

Variável	Média	Desvio Padrão	Coefficiente de Variação
Esperança de Vida (em anos)	66,97	2,39	3,6%
Taxa de Alfabetização (em %)	80,85	9,62	11,9%
PIB per capita ajustado (em dólares ajustados)	4846	1697	35,0%

A variável escolhida deve ser o *PIB per capita ajustado*, pois apresenta a maior heterogeneidade entre as três (maior coeficiente de variação). Uma variável muito homogênea como a esperança de vida, por exemplo, não seria boa escolha, pois se as unidades são muito parecidas quanto à esperança de vida, não há como discriminá-las e, portanto, não há como ordená-las quanto à qualidade de vida.

Solução da Seção 6: Medidas de Posição I

Percentis

Exercício 1:

a)

Mediana:

50% de 100 = 50. A mediana está entre o 50º valor e o 51º valor. Ou seja, a mediana está entre 3,9 e 4,0. A mediana é 3,95. Sendo assim, metade dos chefes de família dessa comunidade tem renda mensal inferior a 3,95 salários-mínimos.

Primeiro Quartil (Q1):

25% de 100 = 25. O primeiro quartil está entre o 25º valor e o 26º valor. Ou seja, o primeiro quartil está entre 2,3 e 2,3. O primeiro quartil é 2,3, ou seja, um quarto dos chefes de família dessa comunidade tem renda mensal inferior a 2,3 salários-mínimos..

Terceiro Quartil (Q3):

75% de 100 = 75. O terceiro quartil está entre o 75º valor e o 76º valor. Ou seja, o terceiro quartil está entre 6,0 e 6,1. O terceiro quartil é 6,05. Sendo assim, três quartos dos chefes de família dessa comunidade têm renda mensal inferior a 6,05 salários-mínimos.

b)

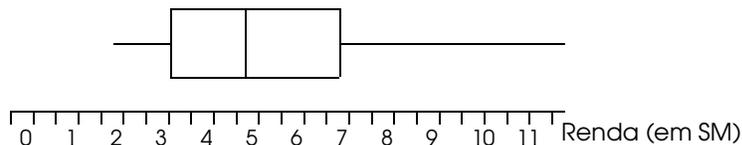
$$DQ = Q3 - Q1 = 6,05 - 2,30 = 3,75$$

$$1,5 \times DQ = 1,5 \times 3,75 = 5,63$$

$$Q1 - 1,5 \times DQ = 2,30 - 5,63 = -3,33$$

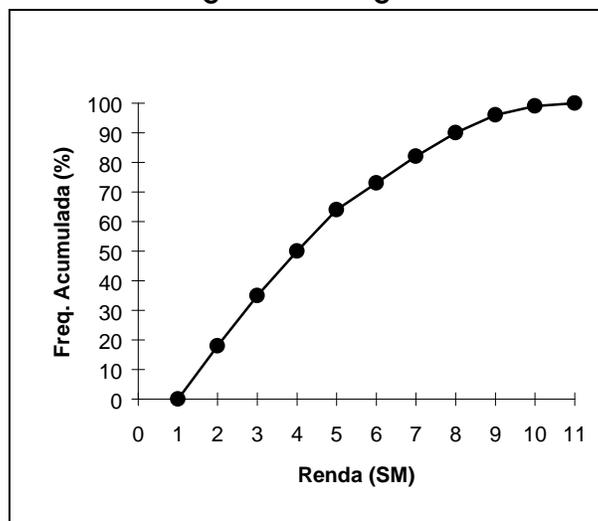
$$Q3 + 1,5 \times DQ = 6,05 + 5,63 = 11,68$$

Figura 6.1.2: Box-Plot.



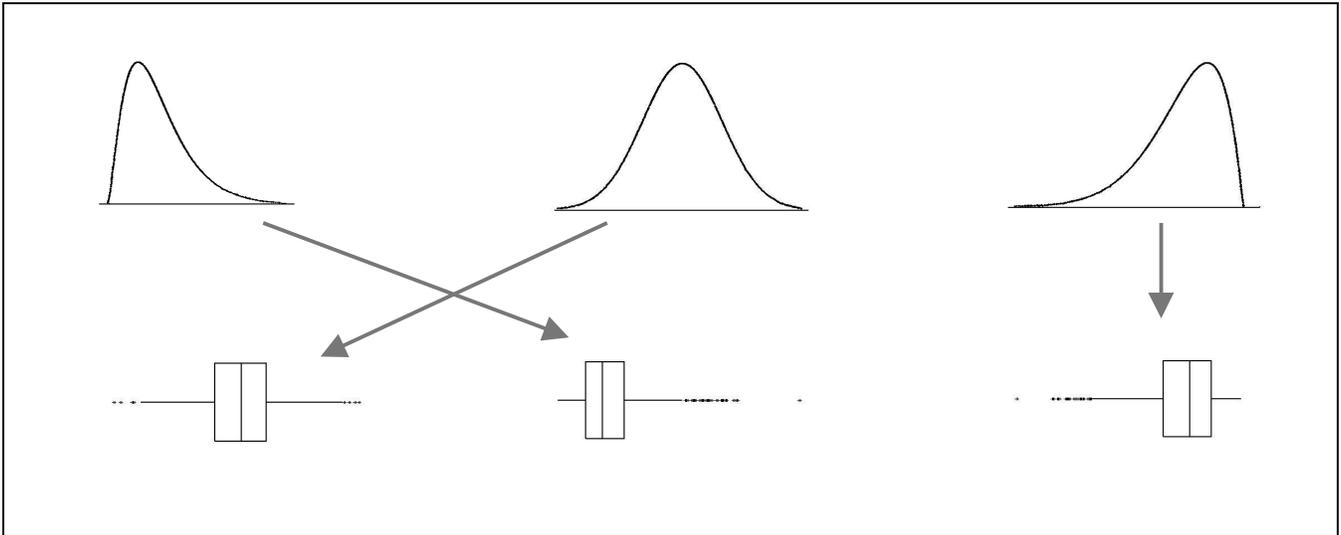
c)

Figura 6.1.3: Ogiva.



Exercício 2:

Figura 6.2.1:



Exercício 3:

Não, pois o box-plot utiliza a mediana como medida de tendência central, que não é adequada para as distribuições bimodais. O box-plot não mostra os dois picos (as duas modas), que são, neste caso, os valores de tendência central adequados.

Exercício 4:

Com as informações fornecidas, podemos concluir que, na empresa A, somente 25% das pessoas ganham mais do R\$1000,00, enquanto, na empresa B, 50% das pessoas ganham mais do que R\$1000,00 (a distribuição dos salários é simétrica em torno da média, assim a média é igual à mediana). Desse modo, a empresa B seria a escolhida.

Solução da Seção 7: Medidas de Posição II

Escores Padronizados

Exercício 1:

a) O Escore Padronizado para cada aluno em cada teste é calculado com a equação

$$EscorePadronizado = \frac{EscoreOriginal - Média}{DesvioPadrão}$$

onde "média" e "desvio padrão" são calculados para cada teste usando os escores originais de todos os 40 alunos. Por exemplo, o escore padronizado no teste de abdominais para o Pedro é :

$$\frac{32 - 30}{6} = 0,3$$

mostrando que Pedro três desvios padrão acima da média de sua turma. E assim, temos:

Quadro 7.1.1:

Teste	Média	Desvio Padrão	Escores Originais			Escores Padronizados		
			Pedro	João	Manuel	Pedro	João	Manuel
Nº de abdominais em 2 min	30	6	32	40	25	0,3	1,7	-0,8
Salto em extensão (centímetros)	155	23	146	140	135	-0,4	-0,7	-0,9
Suspensão braços flexionados (segundos)	50	8	35	70	30	-1,9	2,5	-2,5
Distância percorrida em 12 min (metros)	1829	274	2256	1726	1511	1,6	-0,4	-1,2
Conhecimento desportivo (0 a 100 pts)	75	12	75	95	62	0	1,7	-1,1

b) Devemos comparar os escores originais dos alunos em cada teste. Note que, para todos os testes, quanto maior o escore original melhor é o desempenho do aluno. Por exemplo, no teste dos abdominais, João se saiu melhor porque fez mais abdominais e Manuel teve o pior desempenho porque fez menos abdominais. Note que poderíamos usar também os escores padronizados, mas não necessariamente.

Quadro 7.1.2:

Teste	Aluno com pior desempenho	Aluno com melhor desempenho
Nº de abdominais em 2 min	Manuel	João
Salto em extensão (centímetros)	Manuel	Pedro
Suspensão braços flexionados (segundos)	Manuel	João
Distância percorrida em 12 min (metros)	Manuel	Pedro
Conhecimento desportivo (0 a 100 pts)	Manuel	João

c) Neste caso, apenas os escores padronizados de todos os testes podem ser comparados para cada aluno. Lembre-se de que, para todos os testes, quanto maior o escore original melhor é o desempenho do aluno. Desse modo, para cada aluno, ordenamos os testes (do pior para o melhor desempenho) na ordem crescente do escore padronizado:

Quadro 7.1.3:

Aluno	Ordenação dos testes segundo desempenho do aluno				
	Pior	2º pior	...	2º melhor	Melhor
Pedro	Suspensão	Salto	Conhecimento	Abdominais	Corrida
João	Salto	Corrida	Conhecimento e Abdominais		Suspensão
Manuel	Suspensão	Corrida	Conhecimento	Salto	Abdominais

d) Se desejarmos uma única medida que expresse o desempenho global do aluno (em todos os testes), podemos calcular a média (ou a mediana) de todos os escores padronizados desse aluno. Assim, Manuel teve o pior desempenho global e João o melhor:

Desempenho Global	Pedro	João	Manuel
Média dos Escores Padronizados	-0,1	1,0	-1,3
Mediana dos Escores Padronizados	0	1,7	-1,1

Note que, se houvesse um teste para o qual maior o escore original do aluno, pior foi seu desempenho (por exemplo, tempo gasto para nadar determinada distância), o escore padronizado deste teste deveria entrar com sinal invertido no cálculo do desempenho global.

Exercício 2:

- a) É errado simplesmente somar todas as variáveis porque elas são medidas em unidades diferentes e, conseqüentemente, em escalas diferentes.
- b) O pesquisador deveria calcular os escores padronizados para todas as variáveis e somar estes escores padronizados e não os valores originais.

Note que, para todas as variáveis, quanto mais alto o escore, melhor a posição do bairro em relação aos outros. Caso existisse uma variável para a qual maior o escore do bairro, pior seria sua qualidade de vida (por exemplo, taxa de mortalidade infantil), o valor do escore padronizado dessa variável deveria entrar com sinal negativo no cálculo do índice global.

Exercício 3:

Para que os pesos nesta média ponderada reflitam a heterogeneidade dos resultados de cada teste, devemos usar como pesos o coeficiente de variação (CV) de cada teste, cujos cálculos são mostrados abaixo:

Teste	Média	Desvio Padrão	CV
Nº de abdominais em 2 min	30	6	0,20
Salto em extensão (centímetros)	155	23	0,15
Suspensão braços flexionados (segundos)	50	8	0,16
Distância percorrida em 12 min (metros)	1829	274	0,15
Conhecimento desportivo (0 a 100 pts)	75	12	0,16

Chamando de Z_1, Z_2, Z_3, Z_4 e Z_5 o escore padronizado de cada teste, a fórmula do Índice Global ponderado pelo CV é descrita abaixo:

$$IG = \frac{(0,20 \times Z_1) + (0,15 \times Z_2) + (0,16 \times Z_3) + (0,15 \times Z_4) + (0,16 \times Z_5)}{0,20 + 0,15 + 0,16 + 0,15 + 0,16}$$

Como exemplo, para os alunos Pedro, João e Manuel temos:

$$IG_{\text{Pedro}} = \frac{(0,20 \times 0,3) + (0,15 \times (-0,4)) + (0,16 \times (-1,9)) + (0,15 \times 1,6) + (0,16 \times 0)}{0,82} = -0,08$$

$$IG_{\text{João}} = \frac{(0,20 \times 1,7) + (0,15 \times (-0,7)) + (0,16 \times 2,5) + (0,15 \times (-0,4)) + (0,16 \times 1,7)}{0,82} = 1,03$$

$$IG_{\text{Manoel}} = \frac{(0,20 \times (-0,8)) + (0,15 \times (-0,9)) + (0,16 \times (-2,5)) + (0,15 \times (-1,2)) + (0,16 \times (-1,1))}{0,82} = -1,28$$

Assim, dentre estes três alunos, Manuel teve o pior desempenho global e João, o melhor.

Exercício 4:

- a) O cálculo dos escores padronizados permanece o mesmo, como mostrado no quadro abaixo para os alunos Pedro, João e Manuel:

Quadro 7.4.1: Continuação item (a) do Exercício 1.

Teste	Média	Desvio Padrão	Escore Original			Escore Padronizado		
			Pedro	João	Manuel	Pedro	João	Manuel
Tempo gasto para nadar 100m(segundos)	30	5	22	30	45	-1,6	0	3

Quanto menor o tempo gasto para se nadar uma distância fixa (100 metros, neste caso), melhor é o desempenho do aluno:

Quadro 7.4.2: Continuação item (b) do Exercício 1.

Teste	Aluno com o pior desempenho	Aluno com o melhor desempenho
Tempo gasto para nadar 100m (segundos)	Manuel	Pedro

Na ordenação, para cada aluno, dos testes de seu pior a seu melhor desempenho, deve-se considerar o valor do escore padronizado do teste de natação com o sinal invertido (negativo):

Quadro 7.4.3: Reformulação item (c) do Exercício 1.

Aluno	Ordenação dos testes segundo desempenho do aluno					
	Pior	2º pior	2º melhor	Melhor
Pedro	Suspensão	Salto	Conhecimento	Abdominais	Corrida e Natação	
João	Salto	Corrida	Natação	Abdominais e Conhecimento		Suspensão
Manuel	Natação	Suspensão	Corrida	Conhecimento	Salto	Abdominais

- b) No cálculo do índice de desempenho global de cada aluno, deve-se também inverter o sinal do escore padronizado do teste de natação:

Reformulação item (d) do Exercício 1.

Desempenho Global	Pedro	João	Manuel
Média dos Escores Padronizados	0,2	0,8	-1,6
Mediana dos Escores Padronizados	0,15	0,85	-1,15

Cálculo dos índices globais para Pedro:

$$\text{Média}_{\text{Pedro}} = \frac{(0,3) + (-0,4) + (-1,9) + (1,6) + (0) - (-1,6)}{6} = \frac{1,2}{6} = 0,2$$

Valores dos escores padronizados de Pedro em ordem crescente: -1,9 -0,4 0 0,3 1,6 1,6

Mediana_{Pedro}: média entre o 3º e o 4º valores ordenados = $(0 + 0,3)/2 = 0,3/2 = 0,15$

Cálculo dos índices globais para João:

$$\text{Média}_{\text{João}} = \frac{(1,7) + (-0,7) + (2,5) + (-0,4) + (1,7) - (0)}{6} = \frac{4,8}{6} = 0,8$$

Valores dos escores padronizados de João em ordem crescente: -0,7 -0,4 0 1,7 1,7 2,5

Mediana_{João}: média entre o 3º e o 4º valores ordenados = $(0 + 1,7)/2 = 1,7/2 = 0,85$

Cálculo dos índices globais para Manuel:

$$\text{Média}_{\text{Manuel}} = \frac{(-0,8) + (-0,9) + (-2,5) + (-1,2) + (-1,1) - (3)}{6} = \frac{-9,5}{6} = 1,6$$

Valores dos escores padronizados de Manuel em ordem crescente: -3 -2,5 -1,2 -1,1 -0,9 -0,8

Mediana_{Manuel}: média entre o 3º e o 4º valores ordenados = $(-1,2 - 1,1)/2 = -2,3/2 = -1,15$

Solução da Seção 8: Associação entre Variáveis Qualitativas

Exercício 1:

a)

Tabela 8.1.2:

Sentença Final	Alegação Inicial do Acusado		Total
	Culpado	Inocente	
Enviado à prisão	41%	81%	44%
Mantido em liberdade	59%	19%	56%
Total	100%	100%	100%

Tabela 8.1.3:

Sentença Final	Alegação Inicial do Acusado		Total
	Culpado	Inocente	
Enviado à prisão	87%	13%	100%
Mantido em liberdade	98%	2%	100%
Total	93%	7%	100%

A porcentagem da primeira casela da Tabela 8.1.2 é lida como:

41% dos réus que tinham alegado culpa foram enviados à prisão.

A porcentagem da primeira casela da Tabela 8.1.3 é lida como:

87% dos réus que foram enviados à prisão tinham alegado culpa.

A primeira porcentagem restringe a contagem de réus presos dentre apenas aqueles 956 réus que alegaram culpa; a segunda porcentagem toma como total (100%) os 450 réus que foram enviados à prisão.

Hipótese no enunciado: *réus que se declaram culpados tendem a sofrer penas mais leves do que aqueles que se declaram inocentes.*

Variável explicativa: alegação inicial do acusado (culpado/inocente).

Variável resposta: sentença final do júri (prisão/liberdade).

- b)** Para verificar a hipótese colocada no item (b), deve analisar a Tabela 8.1.2, que mostra a porcentagem de réus enviados à prisão (ou libertados) separadamente entre aqueles que haviam alegado culpa e aqueles que haviam alegado inocência. Ou seja, analisa-se a tabela que tem as porcentagens somando 100% dentro das categorias da variável explicativa.
- c)** A sentença final não é independente da alegação inicial do acusado: aquele que alega inocência tem quase o dobro de probabilidade (81%) de ser enviado à prisão do que o acusado que alega culpa (41%).

Exercício 2:

a)

Tabela 8.2.1:

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	40 (40%)	60 (60%)	100 (100%)
Feminino	60 (60%)	40 (40%)	100 (100%)
Total	100 (50%)	100 (50%)	200 (100%)

b) Sim, existe associação entre o sexo do condenado e a severidade de sua sentença, pois, 60% dos homens receberam sentença severa mas uma porcentagem menor, 40%, das mulheres receberam uma sentença severa. Ou seja, podemos dizer que os homens têm uma probabilidade 1,5 vez maior que as mulheres de receber uma sentença severa.

c)

Tabela 8.2.2: Nenhuma condenação anterior.

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	10 (100%)	0 (0%)	10 (100%)
Feminino	50 (71%)	20 (29%)	70 (100%)
Total	60 (75%)	20 (25%)	80 (100%)

Tabela 8.2.3: Uma ou duas condenações anteriores.

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	12 (60%)	8 (40%)	20 (100%)
Feminino	10 (50%)	10 (50%)	20 (100%)
Total	22 (55%)	18 (45%)	40 (100%)

Tabela 8.2.4: Mais de três condenações anteriores.

Sexo	Sentença		Total
	Moderada	Severa	
Masculino	18 (26%)	52 (74%)	70 (100%)
Feminino	0 (0%)	10 (100%)	10 (100%)
Total	18 (22%)	62 (78%)	80 (100%)

d) Sim, há associação entre o sexo do condenado e a severidade de sua sentença quando extratificamos pelas categorias da variável "número de condenações anteriores", mas na direção contrária: mulheres tendem a ter penas mais severas que os homens dentro de cada categoria de condenação anterior.

Solução da Seção 9: Associação entre Variáveis Quantitativas

Exercício 1:

Diagrama de dispersão 1: associação linear crescente, $r = 0,842$.

Diagrama de dispersão 2: sem associação, $r = 0,141$.

Diagrama de dispersão 3: associação linear decrescente, $r = -0,787$.

Exercício 2:

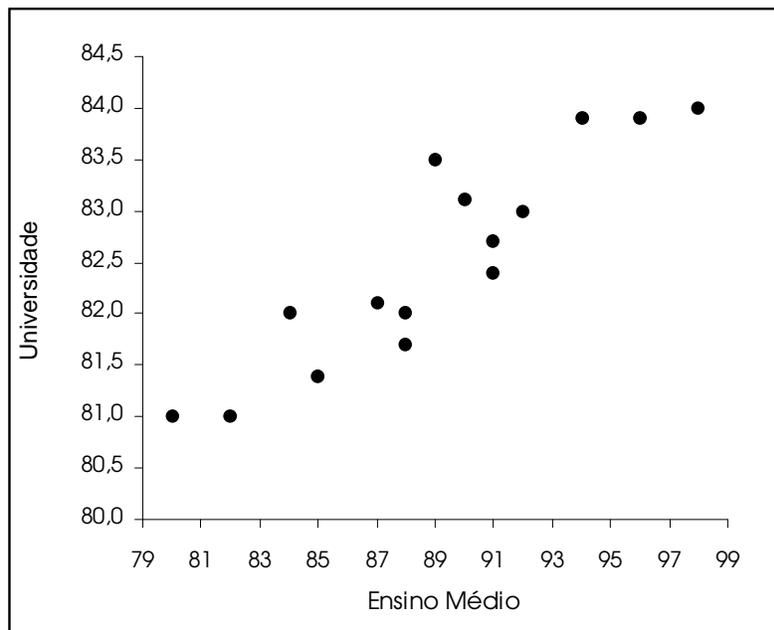
a) Variável explicativa: desempenho do estudante no Ensino Médio.

Variável resposta: desempenho do estudante na Universidade.

Espera-se uma relação crescente: quanto melhor o desempenho do estudante no Ensino Médio melhor seu desempenho na Universidade. Ainda não podemos dizer, antes de olhar o diagrama de dispersão, se a relação é linear e/ou forte.

b)

Figura 9.2.1(a): Diagrama de Dispersão do Desempenho do Aluno na Universidade versus seu Desempenho no Ensino Médio.



O diagrama de dispersão mostra uma forte associação linear crescente (positiva) entre o desempenho do estudante no Ensino Médio e seu desempenho na Universidade: os estudantes que tiveram os melhores (piores) desempenhos no Ensino Médio são, em geral, aqueles com os melhores (piores) desempenhos na Universidade.

c) Após completar os cálculos do Quadro 9.2.1 (a seguir), o cálculo do coeficiente de correlação linear de Pearson é:

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2} \times \sqrt{n(\sum y_i^2) - (\sum y_i)^2}} = \frac{15(110200,5) - (1335,0)(1237,5)}{\sqrt{15(119165,0) - (1335,0)^2} \times \sqrt{15(102107,8) - (1237,5)^2}}$$

$$r = \frac{1653007,5 - 1652062,5}{72,46 \times 14,52} = \frac{945,0}{1052,1} = 0,898$$

O valor do coeficiente de correlação linear de Pearson confirma que é muito forte a correlação positiva entre o desempenho do estudante no Ensino Médio e seu desempenho na Universidade.

Quadro 9.2.1:

l	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	80,0	81,0	6400,0	6561,0	6480,0
2	82,0	81,0	6724,0	6561,0	6642,0
3	84,0	82,1	7056,0	6740,4	6896,4
4	85,0	81,4	7225,0	6626,0	6919,0
5	87,0	82,1	7569,0	6740,4	7142,7
6	88,0	81,7	7744,0	6674,9	7189,6
7	88,0	82,0	7744,0	6724,0	7216,0
8	89,0	83,5	7921,0	6972,3	7431,5
9	90,0	83,1	8100,0	6905,6	7479,0
10	91,0	82,4	8281,0	6789,8	7498,4
11	91,0	82,7	8281,0	6839,3	7525,7
12	92,0	83,0	8464,0	6889,0	7636,0
13	94,0	83,9	8836,0	7039,2	7886,6
14	96,0	83,6	9216,0	6989,0	8025,6
15	98,0	84,0	9604,0	7056,0	8232,0
Soma	1335,0	1237,5	119165,0	102107,8	110200,5

d) Os coeficientes a e b da equação de regressão linear simples são dados por:

$$b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} = \frac{15(110200,5) - (1335,0)(1237,5)}{15(119165,0) - (1335,0)^2} = \frac{1653007,5 - 165206,5}{1787475 - 1782225} = \frac{945,0}{5250,0} = 0,18$$

$$a = \bar{y} - b\bar{x} = \frac{1237,5}{15} - 0,18 \frac{13350}{15} = 82,5 - 0,18(89,0) = 66,48.$$

e) Equação de regressão linear simples relacionando o desempenho do aluno na Universidade ao seu desempenho Ensino Médio:

$$\text{Desemp. Universidade} = 66,48 + 0,18 \times \text{Desemp. Ensino Médio}$$

• Interpretação dos coeficientes da reta:

Interpretação de a = 66,48 ⇒ não tem interpretação prática,, pois não há desempenho no Ensino Médio igual a zero.

Interpretação de b = 0,18 ⇒ a cada um ponto a mais no desempenho no Ensino Médio corresponde, em média, 0,18 pontos a mais no desempenho na Universidade; ou, de modo equivalente, a cada 10 pontos a mais no desempenho no Ensino Médio corresponde, em média, 1,8 pontos a mais no desempenho na Universidade.

• Desenho da equação no Gráfico 1:

Vamos encontrar dois pontos que pertencem à reta, como por exemplo,

$$x = 80 \text{ equivale na reta } y = 66,48 + 0,18 \times (80,0) = 80,88$$

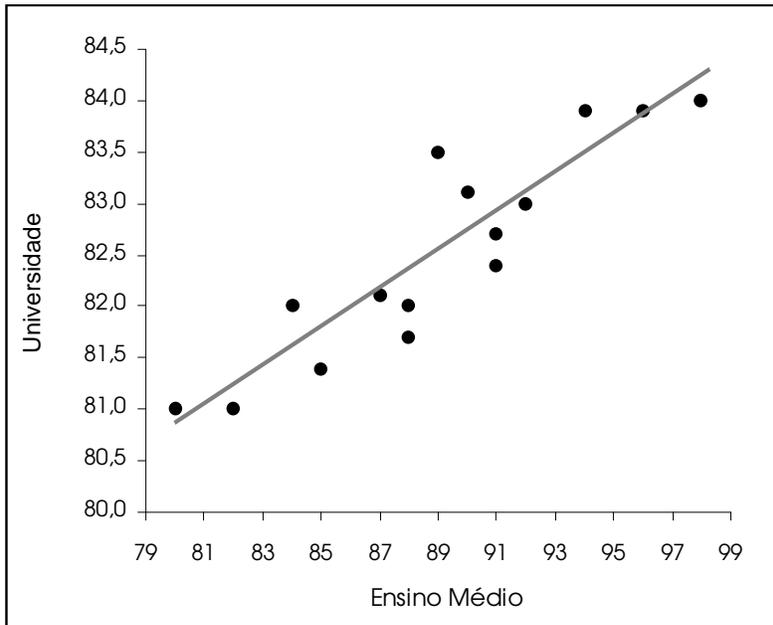
$$x = 98 \text{ equivale na reta } y = 66,48 + 0,18 \times (98) = 84,12$$

Marca-se estes pontos (80 ; 80,88) e (98 ; 84,12) na Figura 9.2.1(b), unindo-os em uma reta.

f) Para um aluno com Desemp. Ensino Médio = 90, temos, de acordo com a equação de regressão, a como estimativa para seu Desemp.Universidade = 66,48 + 0,18 x (90) = 82,68.

A estimativa do desempenho na Universidade para um aluno com desempenho no Ensino Médio igual a 65 não pode ser obtida através desta equação de regressão, pois o valor 65 está fora do intervalo de valores de desempenho no Ensino Médio usado na construção da equação.

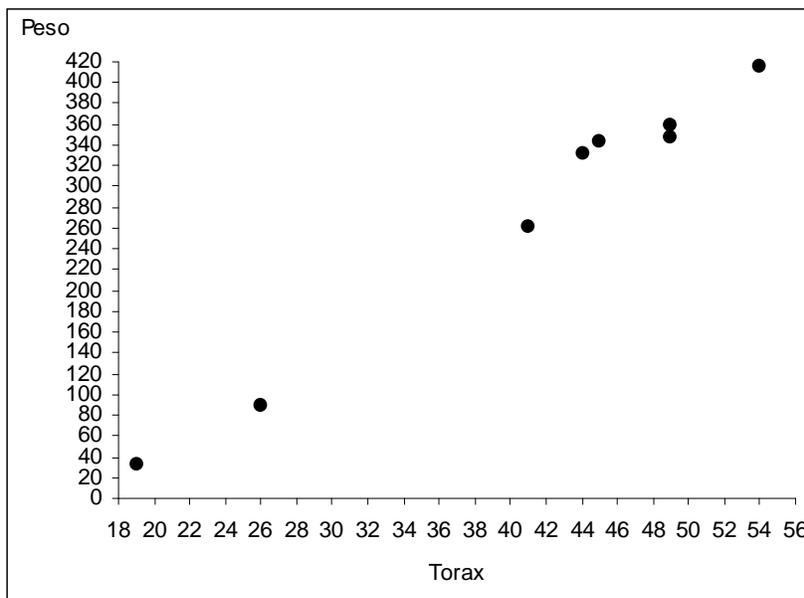
Figura 9.2.1(b): Diagrama de Dispersão do Desempenho do Aluno na Universidade versus seu Desempenho no Ensino Médio e reta de regressão
Desemp. Universidade = 66,48 + 0,18 x Desemp. Ensino Médio.



Exercício 3:

- a) Variável explicativa (X): perímetro do tórax do urso (polegadas).
 Variável resposta (Y): peso do urso (libras)
 Espera-se uma relação crescente: ursos com maior (menor) perímetro do tórax sendo também aqueles com maior (menor) peso.
- d) O diagrama de dispersão mostra uma relação linear crescente (positiva) muito forte entre o perímetro do tórax e o peso do urso macho: quanto maior (menor) o perímetro do tórax, maior (menor) o peso do urso.

Figura 9.3.1(a): Diagrama de Dispersão do o perímetro do tórax versus o peso do urso.



e) Completando o Quadro 9.3.1,

Quadro 9.3.1:

I	x_i	y_i	x_i^2	y_i^2	$x_i y_i$
1	26	90	676	8100	2340
2	45	344	2025	118336	15480
3	54	416	2916	173056	22464
4	49	348	2401	121104	17052
5	41	262	1681	68644	10742
6	49	360	2401	129600	17640
7	44	332	1936	110224	14608
8	19	34	361	1156	646
Soma	327	2186	14397	730220	100972

podemos calcular o coeficiente de correlação linear de Pearson:

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2} \times \sqrt{n(\sum y_i^2) - (\sum y_i)^2}} = \frac{8 \times 100972 - 327 \times 2186}{\sqrt{8 \times 14397 - (327)^2} \times \sqrt{8 \times 730220 - (2186)^2}}$$

$$= \frac{807776 - 714822}{\sqrt{115176 - 106929} \times \sqrt{5841760 - 4778596}} = \frac{92954}{\sqrt{8247} \times \sqrt{1063164}} = \frac{92954}{91 \times 1031} = \frac{92954}{93821} = 0,99$$

cujo valor confirma a muito forte correlação positiva entre o perímetro do tórax e o peso do urso.

f) Os coeficientes a e b da equação de regressão linear simples **Peso = a + b x Tórax** são dados por:

$$b = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{n(\sum x_i^2) - (\sum x_i)^2} = \frac{8 \times 100972 - 327 \times 2186}{8 \times 14397 - (327)^2} = \frac{92954}{8247} = 11$$

$$a = \bar{y} - b\bar{x} = \frac{2186}{8} - 11 \times \frac{327}{8} = 273 - 11 \times 41 = -178$$

g) Equação de regressão linear simples relacionando o peso do urso macho ao perímetro de seu tórax: **Peso = -178 + 11xTorax**

Interpretação de a = -178 ⇒ não tem interpretação prática, pois não existe urso com perímetro de tórax igual a zero.

Interpretação de b = 11 ⇒ a cada polegada a mais no perímetro do tórax corresponde, em média, 11 libras a mais no peso do urso; ou, de modo equivalente, a cada 10 polegadas a mais no perímetro do tórax corresponde, em média, 110 libras a mais no peso do urso.

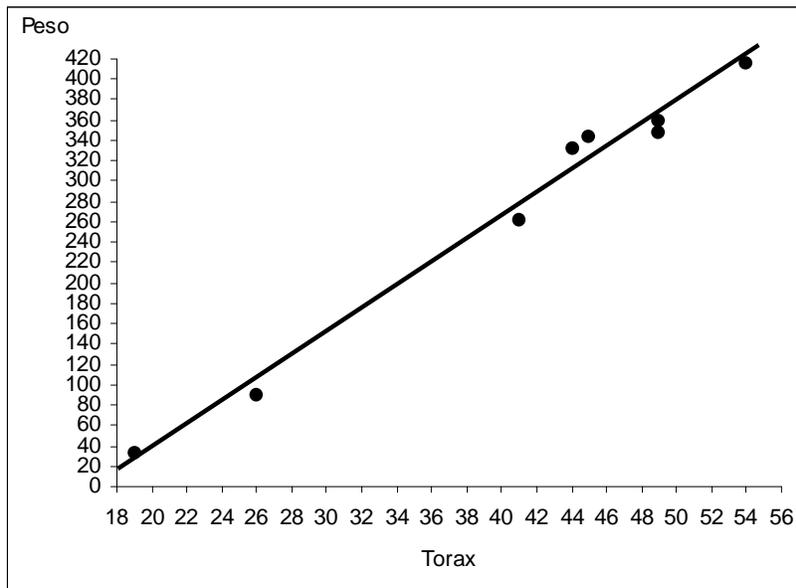
Desenho da equação **Peso = -178 + 11xTorax** na Figura 9.3.1(b) (mostrado a seguir):

Para perímetro do tórax = 18 temos peso = -178 + 11x18 = 20.

Para perímetro do tórax = 54 temos peso = -178 + 11x54 = 416.

h) A estimativa para o peso de um urso que com medida do tórax igual 30 polegadas pode ser obtida pela equação de regressão: tórax = 30 ⇒ **Peso = -178 + 11x30 = 152**.
Para o urso com medida do tórax igual a 70 polegadas, o peso não pode ser estimado pela equação, pois este valor do tórax está fora do intervalo usado para construir a equação.

Figura 9.3.1(b): Diagrama de Dispersão do perímetro do tórax versus o peso do urso
E a reta de regressão linear $\text{Peso} = -178 + 11 \times \text{Torax}$.



Solução da Seção 10: Introdução à Amostragem

Exercício 1:

- a) População amostrada: os 81 atuais senadores brasileiros
Unidade de observação: cada senador
Método de Amostragem: AAS
- b) População amostrada: as 1000 empresas brasileiras que tiveram maiores cotações de ações na *BOVESPA* no ano de 2000
Unidade de observação: cada empresa
Método de Amostragem: AAS
- c) População amostrada: todos os CDs produzidos naquele dia entre 14 e 16 horas
Unidade de observação: cada CD
Método de Amostragem: AS
- d) População amostrada: todos os leitores que deixam esta seção eleitoral entre as 8 e 17 horas
Unidade de observação: cada eleitor
Método de Amostragem: AS
- e) População amostrada: todos os correntistas com aplicações até R\$50.000,00 em 31/12/2000
Unidade de observação: cada correntista
Método de Amostragem: AE
Variável de Estratificação: Valor da aplicação (em reais), definida por três classes:
Primeiro extrato: até R\$5.000,00;
Segundo extrato: entre R\$5.000,01 e R\$20.000,00;
Terceiro extrato: entre R\$20.000,01 e R\$50.000,00.
- f) População amostrada: todos os pacientes de leucemia internados nos hospitais de um estado.
Unidade de observação: cada paciente
Método de Amostragem: AC
Conglomerados: os hospitais

Exercício 2:

A advertência de que, em cada grupo de 10 mil pessoas que fazem o exame, uma apresenta reação alérgica, significa dizer que uma pessoa que fará o exame tem uma probabilidade da grandeza de $1/10.000 = 0,0001$ (ou 0,01%) de ter tal reação alérgica.

Esta probabilidade é constante no tempo, ao contrário do interpretou a paciente. Ela pensou que as primeiras 9.999 pessoas estariam livres da reação alérgica, ao contrário da pessoa número 10 mil, 20 mil, 30 mil, etc.

“Ler” uma probabilidade desta maneira, como “uma em cada n pessoas apresenta tal reação alérgica” não quer significar que todo grupo de 10 mil pessoas tem que ter uma e exatamente uma pessoa com tal reação. Esta é uma média, uma estimativa do número de pessoas que se espera encontrar em cada grupo de n pessoas.

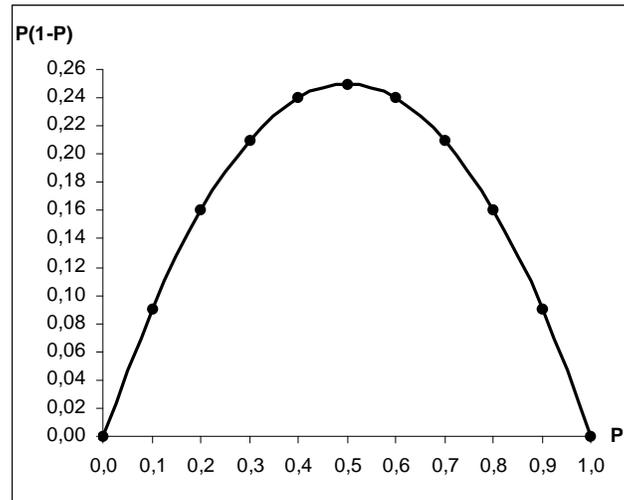
Exercício 3:

a) A Figura 10.3.1 mostra os valores da "heterogeneidade" da variável população, dada por $P(1-P)$, em função dos valores de P . Assim, para completar o gráfico, fazemos os seguintes cálculos:

P	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0
P(1-P)	0,00	0,09	0,16	0,21	0,24	0,25	0,24	0,21	0,16	0,09	0,00

e, assim, a Figura 10.3.1 torna-se:

Figura 10.3.1:



A Figura 10.3.1 mostra que, quanto mais próxima de 0,5 é a probabilidade P , maior é a variabilidade da variável qualitativa binária. Ou seja, quanto mais próxima da divisão exatamente ao meio (50%/50%) em uma característica binária uma população está, maior será a variabilidade desta população.

Assim, para dada característica binária em uma população e dado tamanho de amostra fixo n , a situação em que o erro padrão S atinge seu maior valor é quando $P=0,5$.

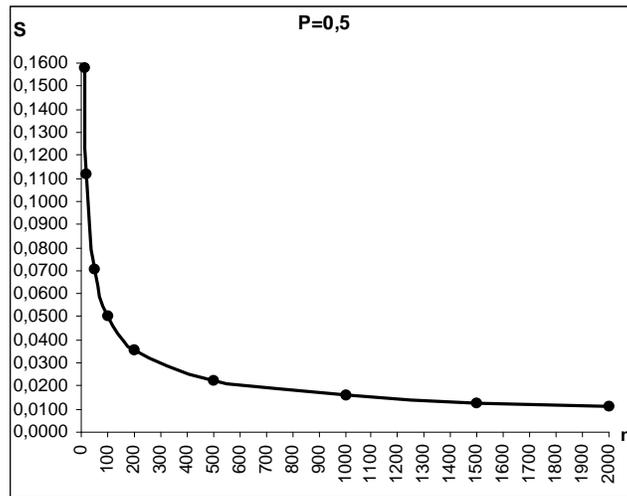
b) Partindo da equação $s = \sqrt{\frac{P(1-P)}{n}}$ com $P=0,5$ temos $s = \sqrt{\frac{0,5(0,5)}{n}} = \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}$.

Assim com $n=10$ temos $s = \frac{1}{2\sqrt{10}} = 0,1581$. Os cálculos de S para os demais valores de n são:

n	10	20	50	100	200	500	1000	1500	2000
s	0,1581	0,1118	0,0707	0,0500	0,0354	0,0224	0,0158	0,0129	0,0112

e, assim, a Figura 10.3.2 torna-se:

Figura 10.3.2:



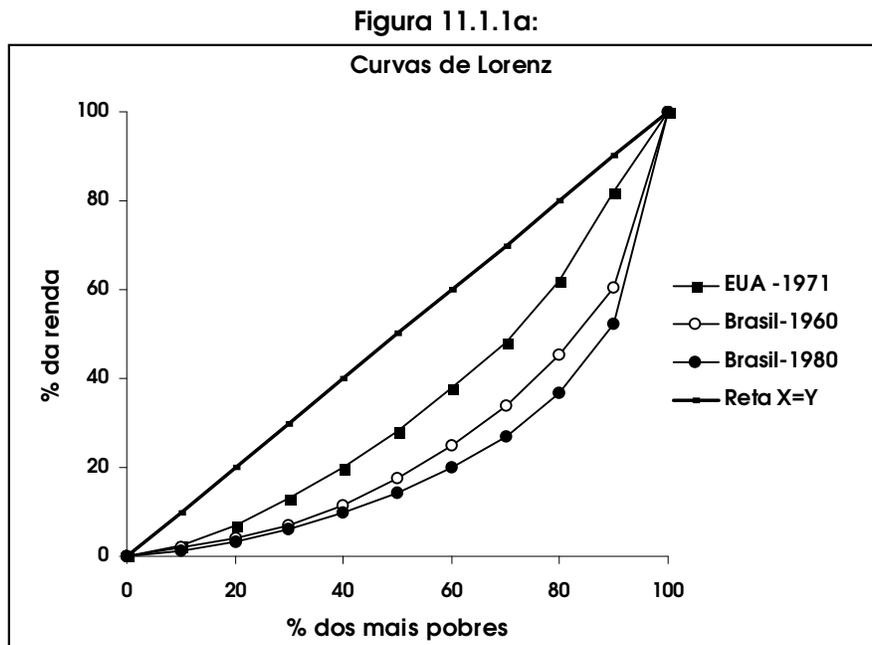
Na Figura 10.3.2 fixamos o valor de $P=0,5$ na fórmula do erro padrão S pois, como vimos no item (a), esta é a situação em que o erro padrão é máximo. Assim, a figura nos mostra que o erro padrão S é inversamente proporcional ao tamanho da amostra n .

Há uma queda muito acentuada no início (até n menores que, digamos, 500) e muito lenta para tamanhos de amostra maiores. Observe, por exemplo, que passar de um tamanho de 100 para 500 reduz o erro padrão de 0,05 para 0,0224 (menos da metade). Entretanto, se aumentarmos também em 400 observações uma amostra que já tenha tamanho 500, o erro padrão reduz menos, para 0,0167 (mais de dois terços).

Solução da Seção 11: Medidas de Desigualdade

Exercício 1:

a)

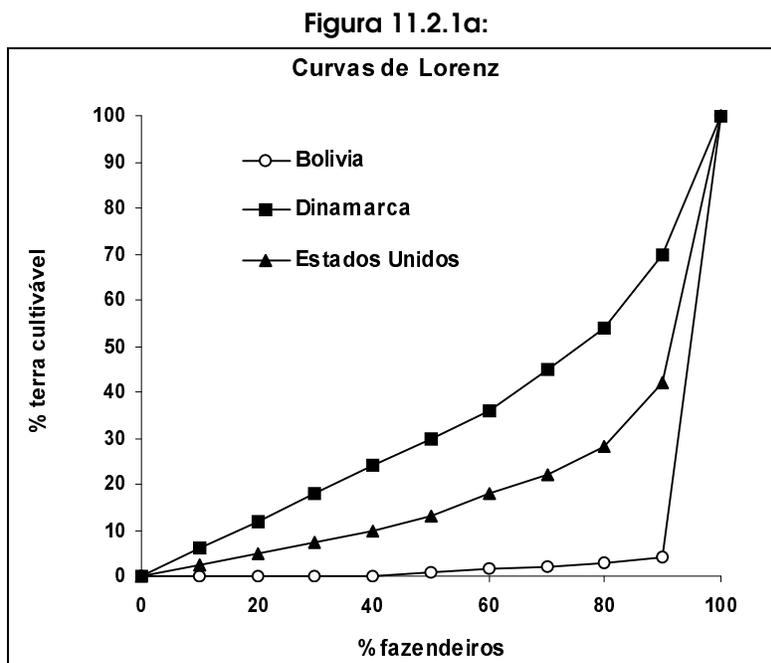


b) O valor do índice de Gini para os dados de renda da população economicamente ativa no Brasil em 1980 é igual a 0,60 , aproximadamente.

c) O Brasil em 1980 apresentava maior desigualdade na distribuição de renda na população economicamente ativa do que em 1960. Os EUA apresentava, em 1971, menor desigualdade na distribuição de renda na população economicamente ativa que o Brasil nas épocas citadas.

Exercício 2:

a)



b) A terra está mais igualmente distribuída entre os fazendeiros na Dinamarca e mais desigual na Bolívia.

Exercício 3:

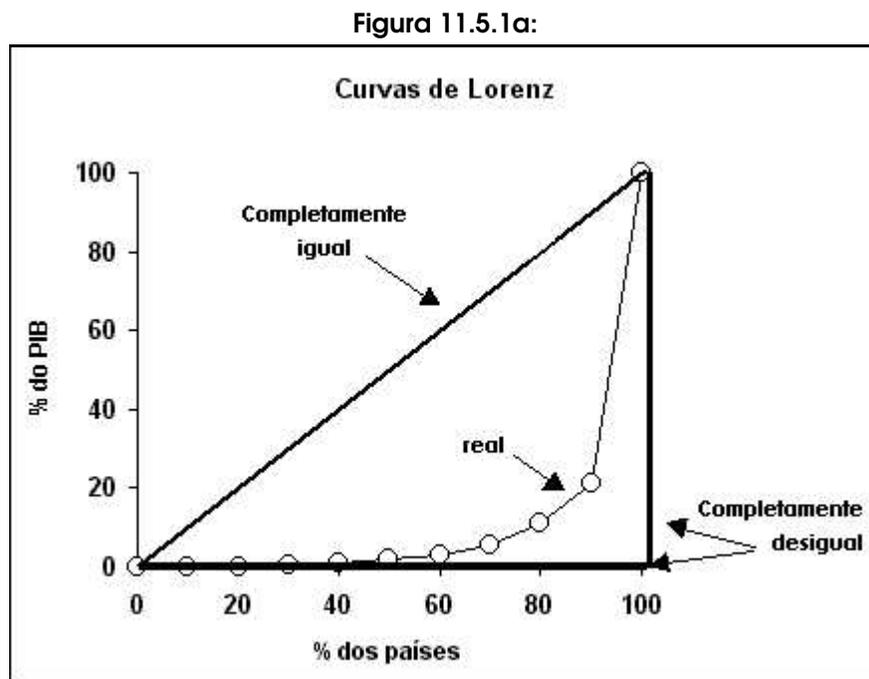
- a) Não, as duas colunas de números são muito parecidas.
- b) As curvas de Lorenz deveriam ser muito similares.
- c) A população total cresce mas mantém o nível de desigualdade de sua distribuição entre os maiores países.

Exercício 4:

Não os menores valores de renda estão no início de curva.
A cada 1% adicional da população no eixo horizontal corresponde uma porcentagem cada vez maior da renda total. Assim, a inclinação da curva é sempre crescente a medida que passamos da esquerda para a direita.

Exercício 5:

- a) e b)



A curva dos dados reais está mais próxima da curva de distribuição *completamente desigual* do PIB do que da curva de distribuição *completamente igual*. Portanto, podemos dizer que há desigualdade na distribuição do PIB entre os países do mundo.

Referências Bibliográficas

- Babbie, E. (1999) *Métodos de Pesquisas de Survey* (tradução). Editora UFMG, 519 pg, ISBN 85-7041-175-8.
- Bolfarine, H e Bussab, W. O. (2000) *Elementos de Amostragem*. Versão preliminar, Instituto de Matemática e Estatística – Universidade de São Paulo, 280 pg.
- COPEVE-UFMG (2001) Home page: <http://www.ufmg.br/copeve/estatisticas/index.html>.
- Harry M. Schey- Department of Biostatistics / University of North Carolina, *THE DISTRIBUTION OF RESOURCES* (1979)
- Huff, D. (1982) *How To Lie With Statistics*. W.W. Norton & Company, 142 pg, ISBN 0-393-31072-8.
- Lopes, P. A. (1999) *Probabilidades e Estatística*. Reichmann & Affonso Editores, 174 pg, ISBN 85-87148-07-9.
- Triola, M. F. (1999) *Introdução à Estatística* (tradução). 7ª edição, Editora LTC, 410 pg, ISBN 85-216-1154-4.
- Zeisel, H. (1985) *Say It With Figures*. 6ª edição, Harper & Row Publishers, 272 pg, ISBN 0-06-181982-4.

Dúvidas, sugestões, críticas ou obtenção de cópias:

**Prof^a Edna Afonso Reis
Departamento de Estatística
Instituto de Ciências Exatas
Universidade Federal de Minas Gerais**

**Caixa Postal 702, CEP 30.123-970
Belo Horizonte, Minas Gerais, Brasil**

**E-mail: edna@est.ufmg.br
HP: www.est.ufmg.br/~edna**