

## LISTA DE EXERCÍCIOS 1 – INE 5113 – PROF. MARCELO MENEZES REIS AMOSTRAGEM E ANÁLISE EXPLORATÓRIA DE DADOS

- 1) Uma população é composta por 2960 elementos que estão ordenados. Se uma amostra sistemática de 20 elementos tivesse que ser retirada desta população, como você procederia?
- 2) Para ser conhecida a opinião dos estudantes da UFSC sobre o Jornal Universitário, foram colhidas as opiniões de 40 estudantes da última fase do curso de jornalismo. O que você tem a comentar sobre este sistema de amostragem?
- 3) Analise as situações descritas abaixo e decida se deve ser realizada uma amostragem ou um censo, justificando:
  - a) Numa linha de produção de empacotamento de café, observar o peso especificado.
  - b) Em uma sala de aula composta por 40 alunos, analisar suas idades.
  - c) Observar se a água de uma lagoa está contaminada.
  - d) Verificar a carga horária diária de trabalho dos funcionários da cozinha de um restaurante
  - e) Num lote de cabos de aço, verificar a resistência dos mesmos à tração.
- 4) Uma amostragem entre os estudantes da computação, engenharia de produção e automação foi realizada da seguinte maneira: considerou-se cada curso como um estrato, e fez-se retirada proporcional. O curso de computação possui 468 alunos, o de engenharia de produção 481 alunos e o curso de automação possui 351 alunos. Do curso de computação foram analisados 27 alunos. Qual o número total de alunos analisados na amostra?
- 5) Numa sala de aula temos 360 homens e 280 mulheres. Faça uma amostragem estratificada proporcional de tamanho 16 considerando o sexo como variável estratificadora. Quantos de cada sexo serão analisados?
- 6) Analise os planos de amostragens apresentados abaixo e se houver incoerências aponte-as.
  - a) Para avaliar o nível sócio econômico dos consumidores de um supermercado, foram entrevistados os consumidores que lá compareceram no 1º sábado do mês, pela tarde.
  - b) Para analisar os laboratórios de pesquisa quanto ao seu investimento na aquisição de tecnologia de ponta, foram enviados questionários e analisadas as respostas daqueles que responderam o questionário.
- 7) Identificar nas pesquisas a seguir quais são as variáveis independentes e as dependentes. JUSTIFIQUE suas respostas.
  - a) Seja uma pesquisa de opinião eleitoral, em que há interesse na preferência declarada de homens e mulheres por candidatos a presidência da república.
  - b) Levantamento do perfil dos alunos dos diferentes cursos do CSE, indagando o curso a que pertence, sexo, idade, se trabalha ou não.
  - c) Pesquisa de mercado sobre um novo refrigerante, procurando identificar a preferência de consumidores de 5 faixas etárias, público masculino e feminino.
  - d) Uma metalúrgica está realizando um experimento, variando a temperatura, percentual de ferro e tempo de fundição para observar a dureza do aço fabricado.
  - e) Avaliação dos salários mensais de professores de universidades públicas e privadas.
  - f) Há uma disciplina de Estatística oferecida para 5 cursos diferentes do CTC da UFSC, e os respectivos colegiados querem saber como os alunos estão se saindo. Há um arquivo de dados com as notas finais dos alunos de cada curso.
  - g) A ONU coletou informações sobre o PIB per capita de 155 países, com o objetivo de obter uma “radiografia da pobreza”. Os países podem ser enquadrados em seis regiões econômico-geográficas.

8) Identificar qual é o grau de mensuração das variáveis descritas a seguir. JUSTIFIQUE suas respostas.

- a) Curso a que pertencem os alunos regularmente matriculados na UFSC em 2012.2.
- b) Preferência declarada de um eleitor por um candidato à presidência da República.
- c) Consumo de combustível de diferentes modelos de automóveis em km/l.
- d) Número de filhos por família em diferentes países.
- e) Número de residentes nos domicílios da grande Florianópolis.
- f) Classificação da redação dos candidatos de um concurso público em “excelente”, “ótima”, “boa”, “regular” e “insuficiente”.
- g) Classificação de pisos cerâmicos produzidos como “perfeitos”, “aceitáveis” e “inaceitáveis”.
- h) Temperatura medida em graus Celsius.
- i) Nível de instrução dos funcionários de uma empresa, medido pelo número de anos completos na escola (excluindo repetências).
- j) Velocidade de automóveis em uma rodovia federal, declarada “legal” ou “em excesso”.

9) Para os casos a seguir indique qual é o método estatístico (tabela, gráfico, medida de síntese) mais apropriado para resumir e interpretar os dados. JUSTIFIQUE suas respostas.

- a) A diretoria de uma empresa especializada em entrega expressa quer ter uma idéia do peso dos pacotes das encomendas, para decidir ou não a adoção de um tipo maior de caixa. Há um arquivo com os pesos em kg (com até 3 casas decimais) das últimas 3000 encomendas. Você tem 5 minutos para fazer uma apresentação para a diretoria que é IGNORANTE em Estatística.
- b) Uma corretora de valores tem 450 clientes, desde pequenos a mega-investidores. A diretoria quer ter uma idéia dos valores investidos por estes clientes, identificando inclusive os “discrepantes”. Você tem 3 minutos para apresentar o relatório, e a diretoria é versada em Análise Exploratória de Dados.
- c) Foram registrados os sexos (masculino e feminino) dos 851 alunos de um curso de graduação da UFSC. O objetivo é apresentar a distribuição da variável sexo para o público em geral.
- d) Estamos avaliando a qualidade de dois processos produtivos de cerâmica, que costumam apresentar comportamentos diferentes. Um aspecto crucial é a granulometria da massa (medida em  $g/m^3$ ). Queremos avaliar qual dos dois processos apresenta maior homogeneidade (ou seja, apresenta menor dispersão), dispomos de 40 medidas de cada um deles, e nós conhecemos análise exploratória de dados.
- e) O Ministério da Saúde está conduzindo um estudo sobre o IDH em 200 municípios de Santa Catarina. Há interesse em fazer uma análise a mais completa possível do IDH, identificando valores típicos e discrepantes. Exige-se uma apresentação gráfica para ser divulgada em um relatório que será lido por técnicos experientes em Estatística.
- f) Uma empresa avícola quer avaliar os pesos (medidos por balanças de precisão) das embalagens de filé de frango, procurando identificar eventuais problemas. Você deve preparar um relatório com os dados de uma amostra de produção, 800 embalagens, que possibilite uma rápida idéia dos pesos, para apresentação aos operadores, “analfabetos em Estatística”.
- g) A direção da montadora quer ter uma idéia, o mais completa possível, da distribuição da renda mensal dos respondentes em função da preferência por modelo. Todos os integrantes da diretoria têm conhecimentos avançados de Estatística, mas exigem que a apresentação seja rápida.
- h) Você precisa apresentar um relatório sobre a dimensão de uma peça. Dispõe de uma amostra de apenas 50 peças, e a diretoria exige uma apresentação rápida e objetiva (sem margem de dúvida se uma disposição diferente poderia alterar as conclusões). A diretoria conhece Estatística.
- i) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos disponíveis para a apresentação, há menos de 100 observações de renda para cada valor de opinião, e a diretoria é versada em Análise Exploratória de Dados.
- j) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o

relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos disponíveis para a apresentação, há menos de 100 observações de renda para cada valor de opinião, e a diretoria é ignorante em Estatística.

k) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos disponíveis para a apresentação, há mais de 100 observações de renda para cada valor de opinião, e a diretoria é versada em Análise Exploratória de Dados.

l) Alguns pesquisadores precisam apresentar um relatório para a diretoria de um banco, mostrando o relacionamento entre as variáveis opinião sobre o banco atual e renda (mostrando a distribuição da renda para cada tipo de opinião) de um grupo de clientes. Há apenas 5 minutos disponíveis para a apresentação, há mais de 100 observações de renda para cada valor de opinião, e a diretoria é ignorante em Estatística.

10) Você é o responsável pela confecção de um relatório sobre o desempenho de vendas das 22 filiais da empresa em que trabalha. A alta direção investiu muito em uma série de promoções para aumentar as vendas, e quer verificar se isso realmente ocorreu: havia sido estabelecido um valor *mínimo aceitável* de 10%, em relação ao mesmo período do ano passado. Os valores de crescimento percentual das vendas estão a seguir:

4.30 6.34 7.38 3.51 10.90 0.28 7.96 8.12 1.58 5.53 4.11 8.43 2.95 8.54  
 18.76 7.34 0.41 8.31 1.54 4.49 4.65 4.19 -1.28 5.67 -0.25 12.00 4.32 1.82  
 -3.15 9.03 2.31 0.17

a) Construa a tabela de frequências que julgar apropriada para resumir os dados. Com base na tabela o que você conclui acerca da eficácia das promoções?

b) Que ferramenta gráfica você utilizaria para resumir os dados? Por quê?

c) Construa o gráfico escolhido no item b, e com base nele, o que você conclui acerca da eficácia das promoções?

d) Que medida(s) de síntese você utilizaria para resumir os dados? Por quê?

e) Calcule a(s) medida(s) escolhida(s) no item d, e com base nela(s), o que você conclui acerca da eficácia das promoções?

11) A padaria que o seu bisavô fundou ao chegar de Portugal, apesar de ainda próspera e com 5 filiais, tem um número talvez excessivo de produtos, o que está aumentando em demasia os custos. Após acalorados debates, sua família decidiu que cinco tipos de pães (aqui chamados de A, B, C, D e E) poderiam ser retirados de produção: mas apenas um deles. Seria arriscado retirar todos, pois a empresa poderia perder a variedade que é sua grande vantagem competitiva. Há interesse na racionalização do processo, mantendo os produtos que têm demanda mais homogênea ao longo do ano. Como você está estudando Estatística neste momento, decidiram que você fará a análise dos dados, para orientar a escolha. Na tabela abaixo você tem as vendas, em milhares de reais, de cada produto nos últimos doze meses.

Pão	Mês 1	Mês 2	Mês 3	Mês 4	Mês 5	Mês 6	Mês 7	Mês 8	Mês 9	Mês10	Mês11	Mês12
A	19.7	15.4	19.3	17.9	19.7	19.5	21.7	19.9	18.6	20.0	17.4	20.5
B	18.7	17.8	18.3	19.4	19.6	16.3	18.4	19.5	19.2	19.7	21.4	20.9
C	20.2	19.8	18.2	19.6	21.3	10.3	12.6	5.5	18.5	22.2	18.7	20.6
D	5.0	4.5	6.7	20.1	19.9	22.4	20.9	19.6	19.6	18.3	8.0	7.0
E	21.2	18.9	19.2	19.6	17.2	19.3	21.9	15.7	20.0	19.3	20.8	18.9

a) Qual a ferramenta gráfica que você utilizaria para descrever os dados? Por quê?

b) Utilizando a ferramenta escolhida no item a, que produto você recomenda retirar de produção?

c) Que medida(s) de síntese você utilizaria para resumir os dados? Por quê?

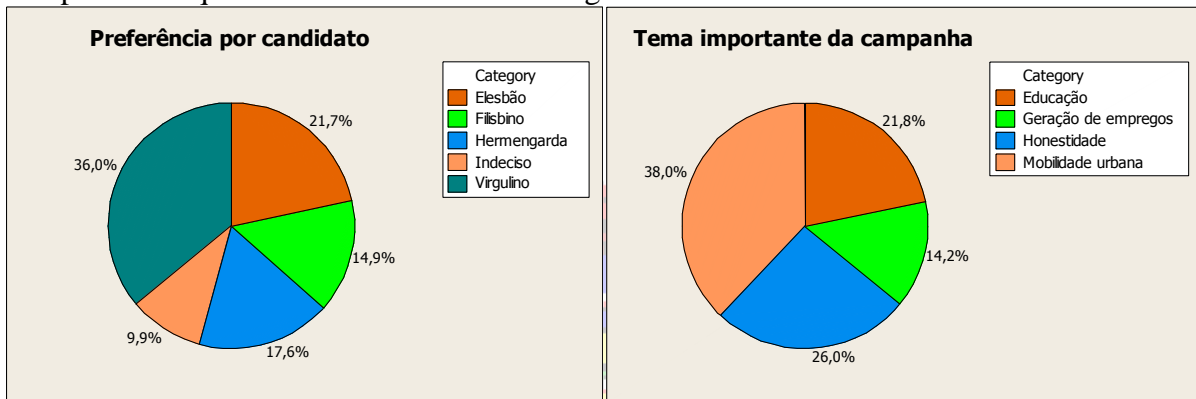
d) Calcule a(s) medida(s) escolhida(s) no item c, e com base nela(s) indique qual produto você recomenda retirar de produção.

12) O instituto de pesquisa ENGABELAX conduziu uma pesquisa de opinião eleitoral para o cargo de prefeito do município de Carambola (300 mil habitantes), apenas com os eleitores de 18 anos ou mais. Foram obtidas as seguintes informações de cada um dos 600 eleitores pesquisados:

- Candidato em quem votaria hoje: Filisbino, Hermengarda, Elesbão, Virgulino ou Indeciso.
- Tema de campanha que o eleitor considerou mais importante para definir seu voto: Geração de empregos, Educação, Mobilidade urbana, Honestidade.
- Região do município onde mora: Sudeste, Nordeste, Noroeste ou Sudoeste.
- Renda familiar do eleitor: em número de salários mínimos vigentes.

Você precisa analisar os resultados da pesquisa e responder algumas perguntas.

Para responder as questões a e b observe os dois gráficos abaixo:

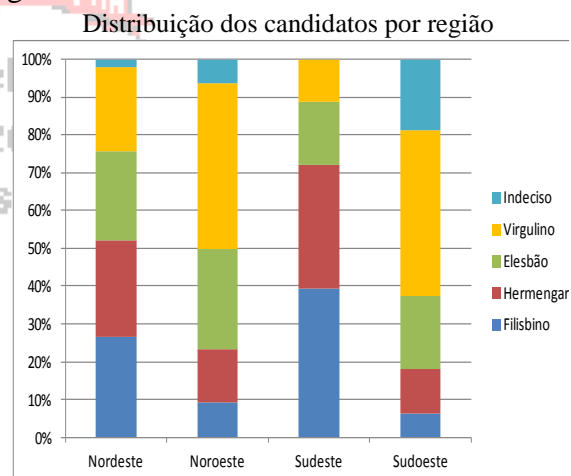


a) Com base no gráfico da esquerda qual é o candidato favorito? JUSTIFIQUE.

b) Qual é o tema mais importante da campanha na opinião dos eleitores? JUSTIFIQUE.

Para responder as questões c e d observe a tabela e o gráfico abaixo:

Candidato	Região				Total geral
	NE	NO	SE	SO	
<b>Filisbino</b>					
Freq.	27	16	31	15	89
% região	26,5%	9,4%	39,2%	6,1%	14,9%
<b>Hermengarda</b>					
Freq.	26	24	26	29	105
% região	25,5%	14,0%	32,9%	11,9%	17,6%
<b>Elesbão</b>					
Freq.	24	45	13	47	129
% região	23,5%	26,3%	16,5%	19,3%	21,6%
<b>Virgulino</b>					
Freq.	23	75	9	107	214
% região	22,5%	43,9%	11,4%	43,9%	35,9%
<b>Indeciso</b>					
Freq.	2	11	0	46	59
% região	2,0%	6,4%	0,0%	18,9%	9,9%
<b>Total Freq.</b>	<b>102</b>	<b>171</b>	<b>79</b>	<b>244</b>	<b>596</b>
<b>Total% região</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

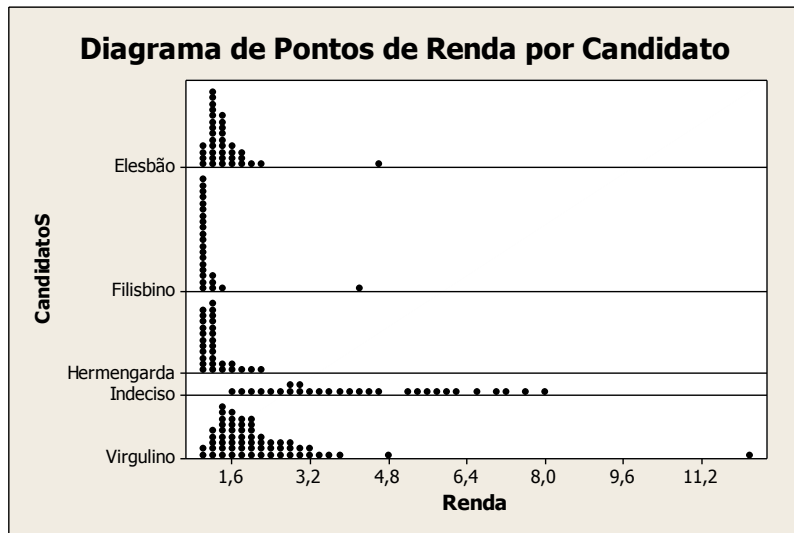


c) As regiões NO e SO são as melhores para o candidato Virgulino? JUSTIFIQUE.

d) Existe relação entre Candidato e Região? JUSTIFIQUE.

Para responder as questões e, f, e g, observem as tabelas e gráficos a seguir:

Candidato	Renda (em salários mínimos)							
	Média	N	D.padrão	Mínimo	Máximo	Q.Inferior	Mediana	Q.Superior
Elesbão	1,38	130	0,37	1,02	4,60	1,17	1,30	1,48
Virgulino	1,99	215	0,95	1,04	12,24	1,46	1,81	2,27
Hermengarda	1,18	105	0,20	1,00	2,22	1,07	1,11	1,22
Filisbino	1,08	89	0,35	1,00	4,28	1,01	1,02	1,05
Indeciso	3,99	59	1,66	1,59	8,04	2,70	3,58	5,12
<b>Total</b>	<b>1,78</b>	<b>598</b>	<b>1,15</b>	<b>1,00</b>	<b>12,24</b>	<b>1,11</b>	<b>1,39</b>	<b>1,94</b>



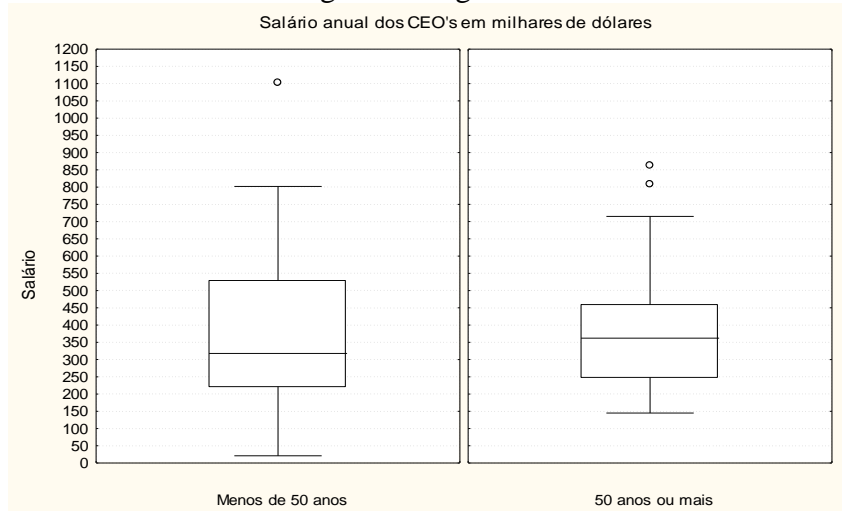
- e) Os eleitores de qual candidato apresentam rendas com tendência central mais elevada? JUSTIFIQUE.
- f) Os eleitores de qual candidato apresentam rendas com maior variabilidade (calcule outras medidas se necessário)? JUSTIFIQUE.
- g) A distribuição das rendas por candidato pode ser considerada assimétrica? JUSTIFIQUE.

13) Os dados a seguir referem-se aos salários anuais (em dólares) dos diretores executivos de 60 “pequenas” empresas americanas (com vendas anuais entre 5 e 350 milhões de dólares). Os executivos foram classificados em dois grupos de acordo com a sua idade: menos de 50 anos, 50 anos ou mais. Responda as questões apresentadas, JUSTIFICANDO suas respostas

Idades	Medidas de síntese para salários anuais (milhares de dólares)		
	Número	Média	D. Padrão
Menos de 50	35	382,2	219,98
50 ou mais	25	398,61	197,92

- a) Com base apenas na tabela acima, qual das faixas de idade tem os maiores salários?
- b) Com base apenas na tabela acima, qual das faixas de idade tem os salários mais homogêneos?

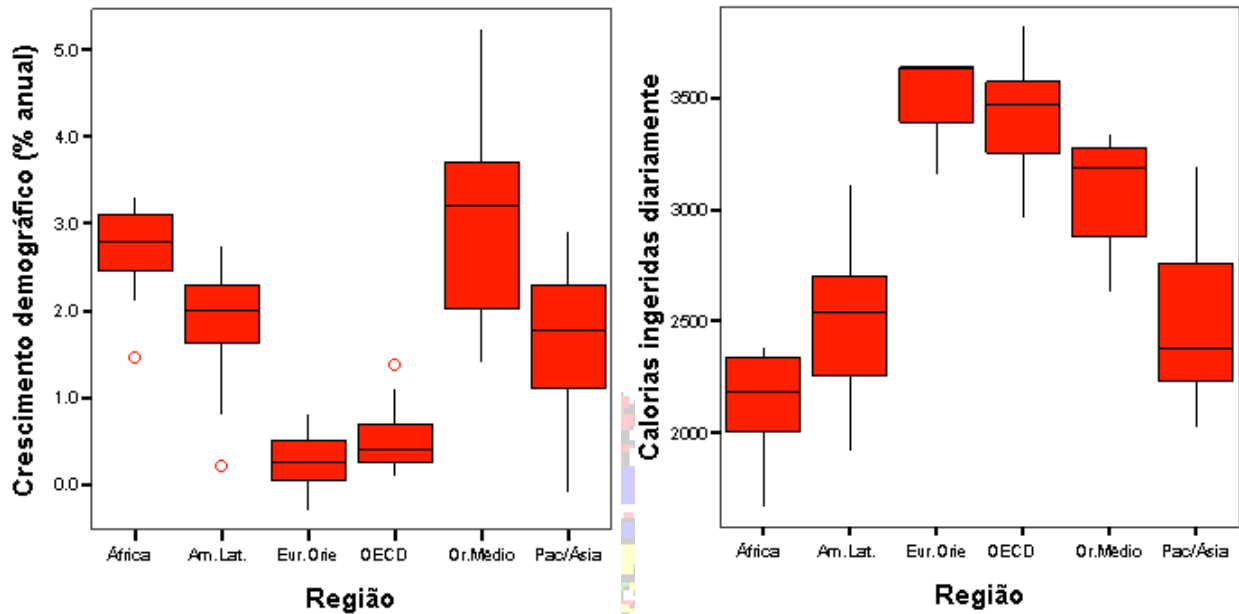
14) Ainda os dados do Exercício 13. A seguir os diagramas em caixa dos salários dos executivos.



- a) Analise os diagramas em caixa quanto ao valor típico, assimetria, dispersão e valores discrepantes.
- b) Com base no diagrama em caixas qual das faixas de idade está ganhando melhor?

15) A ONU realizou uma pesquisa registrando os crescimentos demográficos e médias de calorias diárias ingeridas em vários países. Os países foram agrupados em seis regiões: OECD (EUA,

Canadá, Austrália, Nova Zelândia e Europa Ocidental), África, América Latina, Oriente Médio, Europa Oriental, e Pacífico/Ásia. Os diagramas em caixa das variáveis estão abaixo. Faça a análise dos dois diagramas no que tange aos valores típicos, assimetria, dispersão e valores discrepantes. Qual é a sua opinião sobre a qualidade de vida nestas seis regiões?



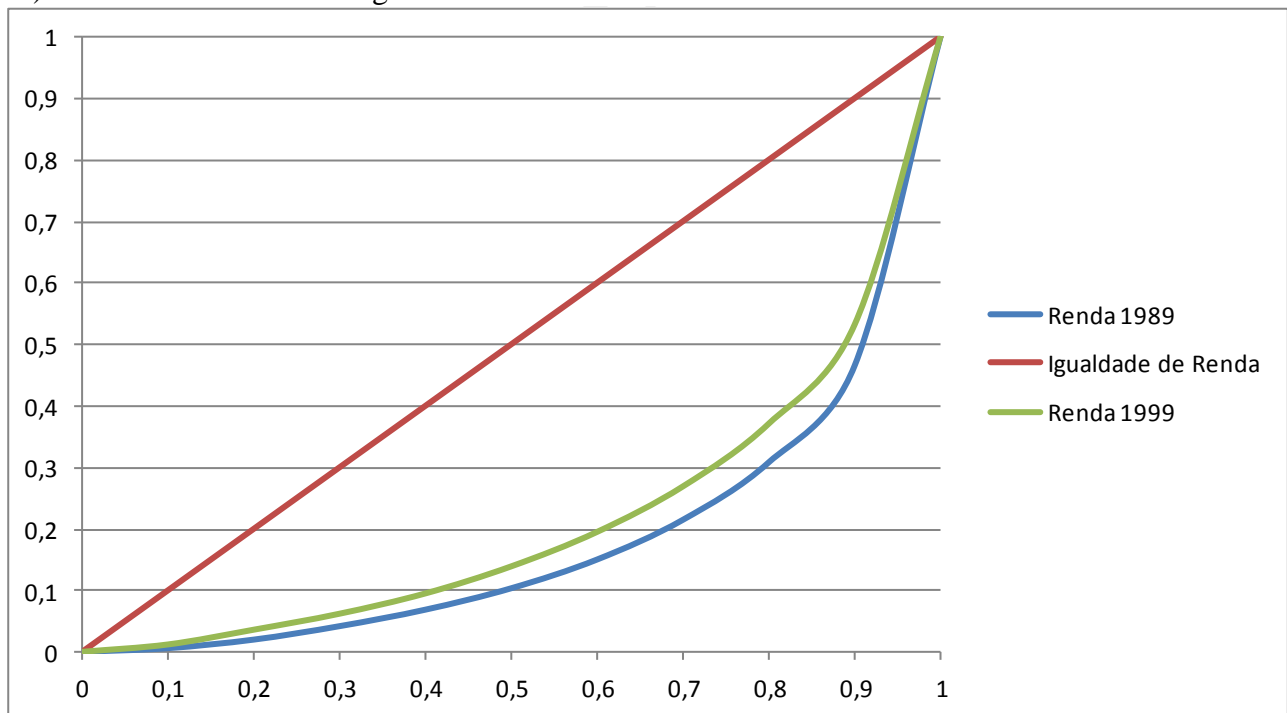
Fonte: ONU, 1995

16) Construa a curva de Lorenz para os dados a seguir:

i	1	2	3	4	5	6	7	8	9	10
Renda	0,6	1,4	2,2	2,7	3,5	4,7	6,5	9,4	15,8	53,8

Parece haver concentração de renda? JUSTIFIQUE.

17) As curvas de Lorenz a seguir mostram a renda familiar no Brasil em 1989 e em 1999.



Fonte: BRAULE, Ricardo. Estatística Aplicada com Excel: para cursos de administração e economia. Rio de Janeiro: Campus, 2001.

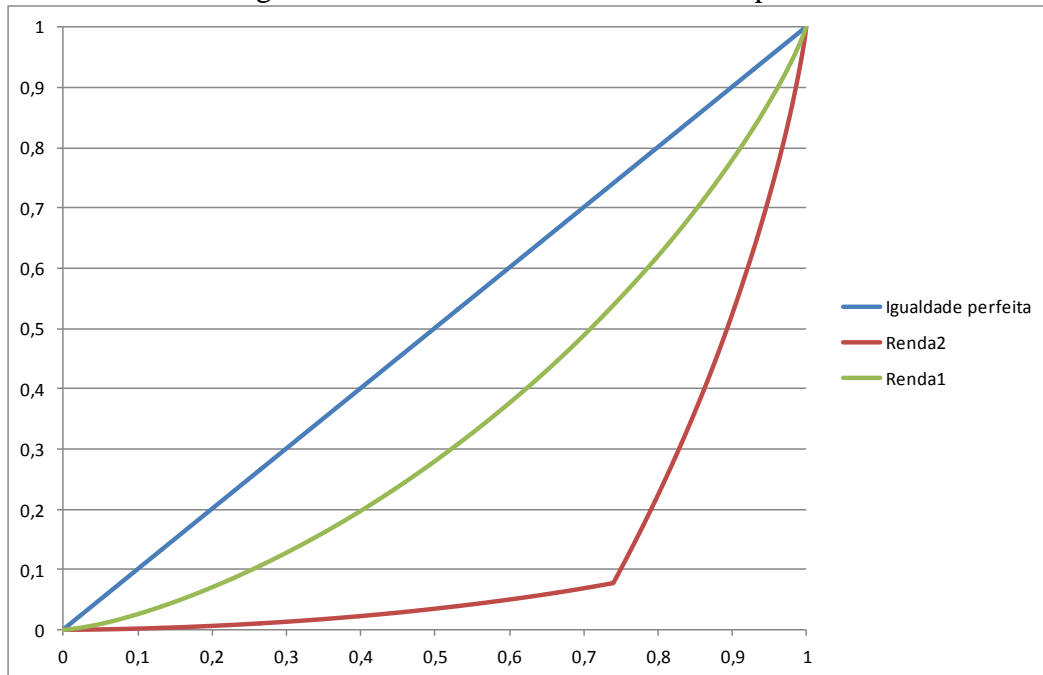
Com base na curva o que se pode dizer da concentração de Renda de 1989 a 1999? JUSTIFIQUE.

18) A série a seguir mostra os índices de Gini da renda familiar no Brasil de 1995 a 2007.

Ano	1995	1996	1997	1998	1999	2001	2002	2003	2004	2005	2006	2007
Índice	0,567	0,568	0,57	0,567	0,56	0,558	0,553	0,545	0,535	0,532	0,528	0,52

Com base na tabela há indício de redução na desigualdade de renda no período analisado? JUSTIFIQUE.

19) As curvas de Lorenz a seguir mostram a renda familiar em dois países.



- a) Um deles apresenta índice de Gini igual a 0,701, e o outro 0,377. Qual é o índice correspondente à Renda1 e qual o correspondente à Renda2? JUSTIFIQUE.
- b) Em qual dos casos ocorre maior desigualdade de renda? JUSTIFIQUE.

Menezes  
Reis