

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

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?

*O primeiro passo é calcular o tamanho do intervalo de retirada dos elementos. Há 2960 elementos na população ( $N = 2960$ ) e a amostra terá 20 elementos ( $n = 20$ ), então:  $k = N/n = 2960/20 = 148$*

*A cada 148 elementos um será retirado para fazer parte da amostra. Para evitar qualquer viesamento da amostra é preciso sortear o ponto de partida: sorteia-se (por meio de qualquer mecanismo honesto) um número entre 1 e 148, e a partir deste, sempre somando-se 148 ao anterior, retiram-se todos os elementos até completar o tamanho da amostra. Se, por exemplo, fosse sorteado o número 21, o elemento correspondente a ele faria parte da amostra, o próximo será o elemento 169 ( $21 + 148$ ), depois o 317 ( $169 + 148$ ), e assim por diante.*

2) Para ser conhecida a opinião dos estudantes da UFSC sobre o Jornal Universitário, foram colhidas as opiniões de 40 estudante da última fase do curso de jornalismo. O que você tem a comentar sobre este sistema de amostragem?

*O plano é totalmente inadequado, se o público alvo é composto por todos os estudantes da UFSC alunos de TODOS os cursos, e de diversas fases deveriam ter sua opinião registrada, e não apenas os da última fase de um curso específico.*

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.

*Como não se trata de um teste destrutivo, supondo que seja viável (há uma balança precisa e a produção não é atrasada em demasia), o ideal seria pesar TODOS os pacotes, em outras palavras realizar um CENSO (prevenindo, inclusive, futuras ações de consumidores insatisfeitos).*

b) Em uma sala de aula composta por 40 alunos, analisar suas idades.

*Como a população é pequena e a característica é de fácil mensuração deve ser realizado um CENSO.*

c) Observar se a água de uma lagoa está contaminada.

*Obviamente não é possível analisar toda a água da lagoa... Deve-se proceder a uma pesquisa por AMOSTRAGEM, coletando amostras em diversos pontos da lagoa, e em diferentes profundidades.*

d) Verificar a carga horária diária de trabalho dos funcionários da cozinha de um restaurante.

*Por necessidades políticas (para evitar reclamações dos funcionários acerca de excesso de trabalho para uns e não para outros), deve ser realizado um CENSO.*

c) Num lote de cabos de aço, verificar a resistência dos mesmos à tração.

*Obviamente não é possível testar a resistência à tração de todos os cabos... Por se tratar de um teste destrutivo deve ser realizada uma pesquisa por AMOSTRAGEM.*

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?

*Se for retirada uma amostra estratificada proporcional o percentual de participação de cada estrato na amostra deve ser igual ao percentual de cada estrato na população. Há um total de 1300 alunos ( $468 + 481 + 351$ ): os alunos de computação correspondem a 36% do total ( $468/1300 = 0,36$ ). Então eles correspondem a 36% da amostra: se são 27 na amostra, a amostra tem um total de 75 elementos ( $27/0,36 = 75$ ).*

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?

*Basta verificar o percentual de homens e mulheres no total. Há 640 pessoas (360 + 280), das quais 360 são homens, perfazendo 56,25% de homens e 43,75% de mulheres. Se a amostra deve ter 16 elementos, 9 serão homens e 7 mulheres.*

6) Analise os planos de amostragem apresentados abaixo e se houver incoerências aponte-as.

a) Para avaliar o nível sócio econômico dos consumidores de um supermercado, analisaram-se os consumidores que compareceram ao supermercado no primeiro sábado do mês, durante o período da tarde.

*O plano é incoerente por três razões:*

*- limita-se a apenas um dia da semana (o sábado), os clientes que freqüentam o estabelecimento em outros dias não serão pesquisados;*

*- a pesquisa será realizada apenas no período vespertino, aqueles que vão ao supermercado pela manhã ou à noite não serão pesquisados;*

*- restringe à pesquisa a apenas um dia do mês, o que deixará de fora uma grande quantidade de consumidores.*

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.

*Esta metodologia não tem validade estatística, pois se pode contar apenas com aqueles que tiveram a boa vontade de responder o questionário, o que pode ser apenas uma pequena fração total, e cujos comentários podem levar a uma compreensão equivocada dos investimentos em pesquisa.*

7) Identificar nas pesquisas a seguir quais são as variáveis independentes e as dependentes. JUSTIFIQUE suas respostas.

a) *Variável independente: sexo do respondente. Variável dependente: preferência declarada. O sexo do respondente PODE influenciar a sua preferência eleitoral.*

b) *Variável independente: curso do CSE. Variáveis dependentes: sexo, idade, situação laboral. O levantamento pretende caracterizar os alunos por curso, que então PODE influenciar os valores das variáveis sexo, idade e situação laboral.*

c) *Variáveis independentes: sexo e faixa etária do consumidor. Variável dependente: preferência pelo refrigerante. As variáveis sexo e faixa etária do consumidor PODEM influenciar a preferência pelo refrigerante.*

d) *Variáveis independentes: temperatura, percentual de ferro e tempo de fundição. Variável dependente: dureza do aço. A variação dos valores de percentual de ferro, tempo de fundição e temperatura PODERÃO influenciar diretamente os valores da dureza do aço.*

e) *Variável independente: tipo de universidade (pública ou privada). Variável dependente: salário dos professores. O tipo de universidade PODE influenciar nos valores dos salários dos professores.*

f) *Variável independente: curso do CTC. Variável dependente: notas finais em Estatística. O fato do aluno estar no curso A, B, ou C (com maior ou menor carga horária de matemática) PODE influenciar a sua nota na disciplina de Estatística.*

g) *Variável independente: regiões econômico-geográficas. Variável dependente: PIB per capita. O fato de um país estar na região A, B, ou C (com maior ou menor desenvolvimento) PODE influenciar o valor do PIB.*

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

a) *Variável qualitativa nominal. As realizações da variável curso são atributos (categorias), e não podem ser ordenadas.*

b) *Variável qualitativa nominal. As realizações da variável preferência declarada são atributos*

(categorias), e não podem ser ordenadas.

c) Variável quantitativa contínua. As realizações da variável consumo em km/l são números, e podem assumir (teoricamente) uma infinidade de valores.

d) Variável quantitativa discreta. As realizações da variável número de filhos são números, e podem assumir apenas alguns valores (números inteiros).

e) Variável quantitativa discreta. As realizações da variável número de residentes são números, e podem assumir apenas alguns valores (números inteiros).

f) Variável qualitativa ordinal. As realizações da variável são atributos (categorias), e podem ser ordenadas.

g) Variável qualitativa ordinal. As realizações da variável são atributos (categorias), e podem ser ordenadas.

h) Variável quantitativa contínua. As realizações da variável temperatura em graus Celsius são números, e podem assumir (teoricamente) uma infinidade de valores.

i) Variável quantitativa discreta. As realizações da variável nível de instrução (em número de anos completos) são números, e podem assumir apenas alguns valores (números inteiros).

j) Variável qualitativa ordinal. As realizações da variável velocidade (neste caso) são atributos (categorias), e podem ser ordenadas.

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) Há apenas uma variável envolvida. Esta variável (peso) pode ser considerada quantitativa contínua (por causa da precisão das medidas). O objetivo do estudo é ter uma idéia do peso dos pacotes, o que poderia ser interpretado como descrição da tendência central. O conjunto apresenta mais de 100 dados (3000), há pouco tempo para a apresentação (apenas 5 minutos), e o público não conhece Estatística. Por essas razões pode ser construído um histograma agrupado em classes para os pesos: podemos observar onde está o valor do tipo maior de caixa na escala horizontal e verificar se há uma grande quantidade de encomendas com peso acima dele.

b) Há apenas uma variável envolvida. Esta variável (valores investidos) pode ser considerada quantitativa contínua (suas realizações são números que podem assumir uma infinidade de valores, por causa das grandes diferenças entre os clientes, alguns aplicando dezenas de reais e outros dezenas de milhões). O objetivo do estudo é ter uma idéia dos valores investidos inclusive dos discrepantes, o que pode ser interpretado como uma descrição completa. O conjunto apresenta mais de 100 dados (450), há pouco tempo para a apresentação (apenas 3 minutos), e o público conhece Estatística. Por essas razões pode ser construído um diagrama em caixas para os valores investidos pelos clientes: pode-se observar a tendência central, a dispersão e os valores discrepantes em um diagrama único.

c) Há apenas uma variável envolvida. Esta variável (sexo) é qualitativa nominal (pois assume apenas 2 valores, que são atributos e não podem ser ordenados). O objetivo do estudo é apresentar a distribuição da variável sexo, o que poderia ser interpretado como “descrição de tendência central”. O conjunto apresenta mais de 100 dados (851), não há restrições de tempo para a apresentação, e como se trata do público em geral, presume-se que não conheça Estatística. Por essas razões há várias opções possíveis: tabela de frequências da variável sexo, gráfico em barras ou gráfico em setores para a variável sexo. Lembrando que a apresentação gráfica de resultados costuma facilitar a apreensão da informação por parte do público alvo, o que talvez recomendasse um dos dois gráficos citados.

d) Há duas variáveis envolvidas: processo e granulometria. Processo pode assumir dois valores, e será considerada independente, pois se julga que ela pode influenciar os valores de granulometria, que seria a dependente. Processo é qualitativa nominal, e granulometria é quantitativa contínua, pois suas realizações são números que podem assumir uma infinidade de valores. O objetivo do estudo é identificar qual dos dois processos é o mais homogêneo, o que significa a descrição de tendência central e dispersão da granulometria por processo, para que seja possível a comparação. O conjunto apresenta menos de 100 dados (40), não há restrições de tempo para a

apresentação, e nós conhecemos *Análise Exploratória de Dados*. Por essas razões recomenda-se o cálculo de média, desvio padrão e coeficiente de variação da granulometria nos dois processos: aquele que apresentar o menor coeficiente de variação será o mais homogêneo.

e) Há apenas uma variável envolvida. Esta variável (IDH) pode considerada quantitativa contínua (pois suas realizações são números, que, por resultarem de operações de divisões, podem assumir uma infinidade de valores). O objetivo do estudo é fazer uma análise a mais completa possível, identificando valores típicos e discrepantes, o que pode ser interpretado como uma descrição completa. O conjunto apresenta mais de 100 dados (200), exige-se uma apresentação gráfica, e o público conhece Estatística. Por essas razões pode ser construído um diagrama em caixas para os IDHs: descreve-se tendência central, a dispersão e os valores discrepantes em um diagrama único.

f) Há apenas uma variável envolvida. Esta variável (peso das embalagens) pode considerada quantitativa contínua (pois suas realizações são números, que podem assumir uma infinidade de valores, visto que é utilizada uma balança de precisão). O objetivo do estudo é ter uma rápida ideia dos pesos, procurando identificar problemas: os problemas podem ser pesos muito pequenos ou muito grandes, ou com assimetria, o que caracteriza uma descrição completa. O conjunto apresenta mais de 100 dados (800), exige-se uma apresentação rápida (o que significaria o uso de algum gráfico), e o público não conhece Estatística. Por essas razões pode ser construído um histograma agrupado em classes, o que evidenciaria a quantidade de embalagens acima e abaixo dos limites e uma eventual assimetria nos pesos.

g) Há duas variáveis envolvidas: preferência por modelo e renda mensal. Preferência pode assumir vários valores, e será considerada independente, pois se pretende usá-la como variável de agrupamento, em função dos seus valores serão avaliados os valores da renda mensal, que será a dependente. Preferência por modelo é qualitativa nominal (seria apenas o modelo de automóvel), Renda mensal pode ser considerada quantitativa contínua, especialmente se for medida em salários mínimos (mas não há informação a respeito), pois suas realizações são números e podem assumir uma infinidade de valores. O objetivo do estudo é ter uma ideia a mais completa possível da distribuição de renda por preferência, o que configura uma descrição completa. Não há informações sobre o tamanho do conjunto de dados, mas podemos supor que, sendo uma pesquisa patrocinada por uma montadora, haja mais de 100 observações. A apresentação precisa ser feita de forma rápida, e a diretoria é versada em Estatística. Por essas razões recomenda-se a construção de um diagrama em caixas múltiplo, um diagrama para a renda de cada modelo preferido, sendo todos colocados em um gráfico com a mesma escala. Se o número de modelos for muito grande, podemos agrupá-los em categorias, criando uma nova variável qualitativa: por exemplo, montadora FIAT, Palio, Siena, e Weekend seriam “pequenos”, Punto e Estilo, “médios”, Dobló e Linea “grandes”, reduzindo então o número de diagramas e facilitando a análise.

h) Há apenas uma variável envolvida: dimensão da peça. É razoável imaginar que se disponha de equipamentos de medição razoavelmente precisos, o que permitiria considerar a dimensão como quantitativa contínua, pois suas realizações são números que podem assumir uma infinidade de valores. O objetivo é apresentar um relatório sobre a dimensão da peça, sendo que a técnica escolhida não pode dar margem a interpretações conflitantes se suas características forem modificadas. Exige-se que a apresentação seja rápida, e o público alvo conhece Estatística. Por essas razões recomenda-se a construção de um diagrama de pontos para a dimensão da peça, que permite a visualização do comportamento completo dos dados, e é único para um determinado conjunto de dados.

i) Há duas variáveis envolvidas: opinião sobre o banco atual e renda. Opinião será considerada independente, pois se pretende usá-la como variável de agrupamento, em função dos seus valores serão avaliados os valores da renda dos clientes, que será a dependente. Opinião é qualitativa ordinal (poderia ser classificada como “ótima”, “ruim”, etc.), renda pode ser considerada quantitativa contínua, especialmente se for medida em salários mínimos (mas não há informação a respeito), pois suas realizações são números e podem assumir uma infinidade de valores. O objetivo do estudo é apresentar o relacionamento entre as duas variáveis: não há obrigatoriedade de uma descrição completa, mas enriqueceria mais os resultados. O conjunto tem MENOS de 100

observações, há apenas 5 minutos para realizar a apresentação e a diretoria do banco conhece Estatística. Por essas razões recomenda-se a construção de um diagrama de pontos múltiplo, um diagrama para a renda de cada valor de opinião, sendo todos colocados em um gráfico com a mesma escala.

j) Há duas variáveis envolvidas: opinião sobre o banco atual e renda. Opinião será considerada independente, pois se pretende usá-la como variável de agrupamento, em função dos seus valores serão avaliados os valores da renda dos clientes, que será a dependente. Opinião é qualitativa ordinal (poderia ser classificada como “ótima”, “ruim”, etc.), renda pode ser considerada quantitativa contínua, especialmente se for medida em salários mínimos (mas não há informação a respeito), pois suas realizações são números e podem assumir uma infinidade de valores. O objetivo do estudo é apresentar o relacionamento entre as duas variáveis: não há obrigatoriedade de uma descrição completa, mas enriqueceria mais os resultados. O conjunto tem MENOS de 100 observações, há apenas 5 minutos para realizar a apresentação e a diretoria do banco NÃO conhece Estatística. Por essas razões recomenda-se calcular média e intervalo para a renda de cada valor de opinião sendo todos colocados em um gráfico com a mesma escala: atenção, se o número de observações for muito pequeno, menos de 30, por exemplo, talvez simplesmente apresentar os dados em ordem crescente (rol) fosse mais apropriado.

k) Há duas variáveis envolvidas: opinião sobre o banco atual e renda. Opinião será considerada independente, pois se pretende usá-la como variável de agrupamento, em função dos seus valores serão avaliados os valores da renda dos clientes, que será a dependente. Opinião é qualitativa ordinal (poderia ser classificada como “ótima”, “ruim”, etc.), renda pode ser considerada quantitativa contínua, especialmente se for medida em salários mínimos (mas não há informação a respeito), pois suas realizações são números e podem assumir uma infinidade de valores. O objetivo do estudo é apresentar o relacionamento entre as duas variáveis: não há obrigatoriedade de uma descrição completa, mas enriqueceria mais os resultados. O conjunto tem MAIS de 100 observações, há apenas 5 minutos para realizar a apresentação e a diretoria do banco conhece Estatística. Por essas razões recomenda-se a construção de um diagrama em caixas múltiplo, um diagrama para a renda de cada valor de opinião, sendo todos colocados em um gráfico com a mesma escala.

l) Há duas variáveis envolvidas: opinião sobre o banco atual e renda. Opinião será considerada independente, pois se pretende usá-la como variável de agrupamento, em função dos seus valores serão avaliados os valores da renda dos clientes, que será a dependente. Opinião é qualitativa ordinal (poderia ser classificada como “ótima”, “ruim”, etc.), renda pode ser considerada quantitativa contínua, especialmente se for medida em salários mínimos (mas não há informação a respeito), pois suas realizações são números e podem assumir uma infinidade de valores. O objetivo do estudo é apresentar o relacionamento entre as duas variáveis: não há obrigatoriedade de uma descrição completa, mas enriqueceria mais os resultados. O conjunto tem MAIS de 100 observações, há apenas 5 minutos para realizar a apresentação e a diretoria do banco NÃO conhece Estatística. Por essas razões recomenda-se a construção de um histograma agrupado em classes múltiplo, um histograma para a renda de cada valor de opinião, sendo todos colocados em um gráfico com a mesma escala.

10) a) Uma tabela de frequências agrupada em classes é mais apropriada, porque a variável é quantitativa e apresenta grande variação nos valores (pode ser considerada contínua).

Classes	Frequência	Ponto Médio
-3,15  -- 0,51	6	-1.32
0,51  -- 4,17	7	2.34
4,17  -- 7,83	10	6.00
7,83  -- 11,49	7	9.66
11,49  -- 15,15	1	13.32
15,15  -- 18,81	1	16.98
Total	32	-

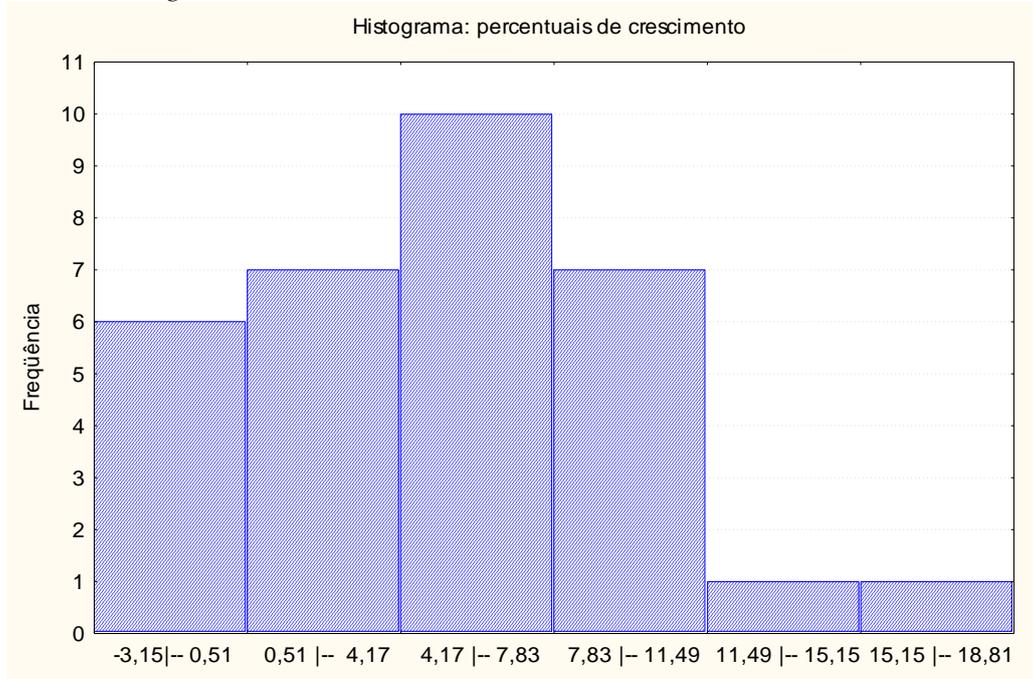
A esmagadora maioria dos percentuais de crescimento está abaixo de 10% (23 filiais, 71,875%). Conclui-se que a promoção não obteve o resultado esperado.

b) Poderia utilizar duas ferramentas:

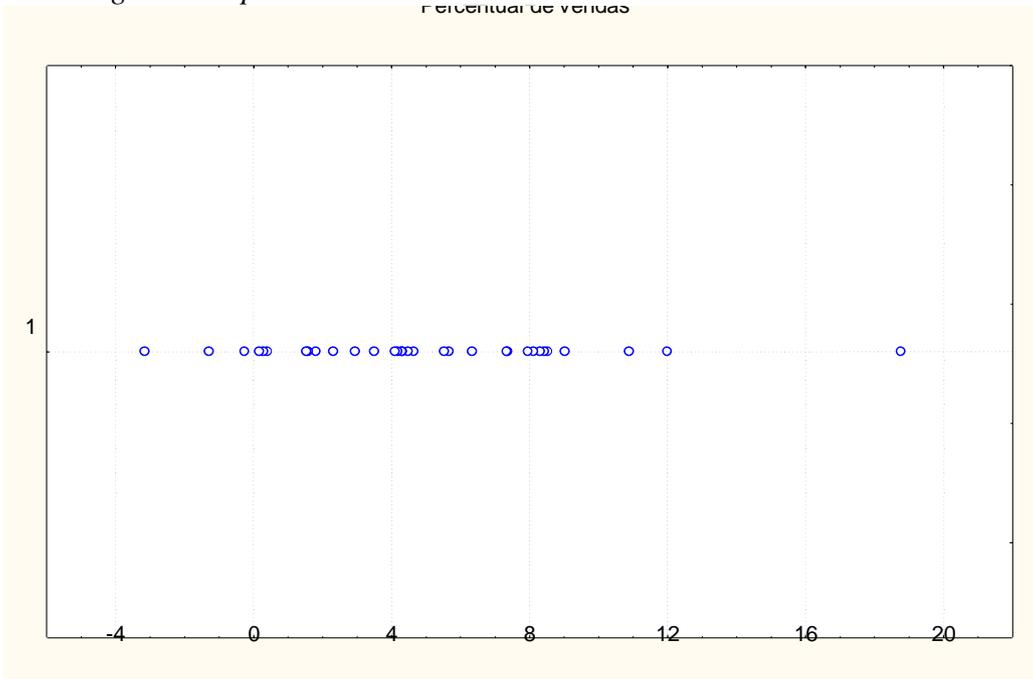
b.1 – Histograma para dados agrupados em classes: porque a variável é quantitativa e apresenta grande variação nos valores (pode ser considerada contínua), e há uma tabela agrupada em classes disponível.

b.2 – Diagrama de pontos: porque a variável é quantitativa e apresenta grande variação nos valores (pode ser considerada contínua) e trata-se de um conjunto pequeno de dados.

c) Construindo o histograma:



Construindo o diagrama de pontos:



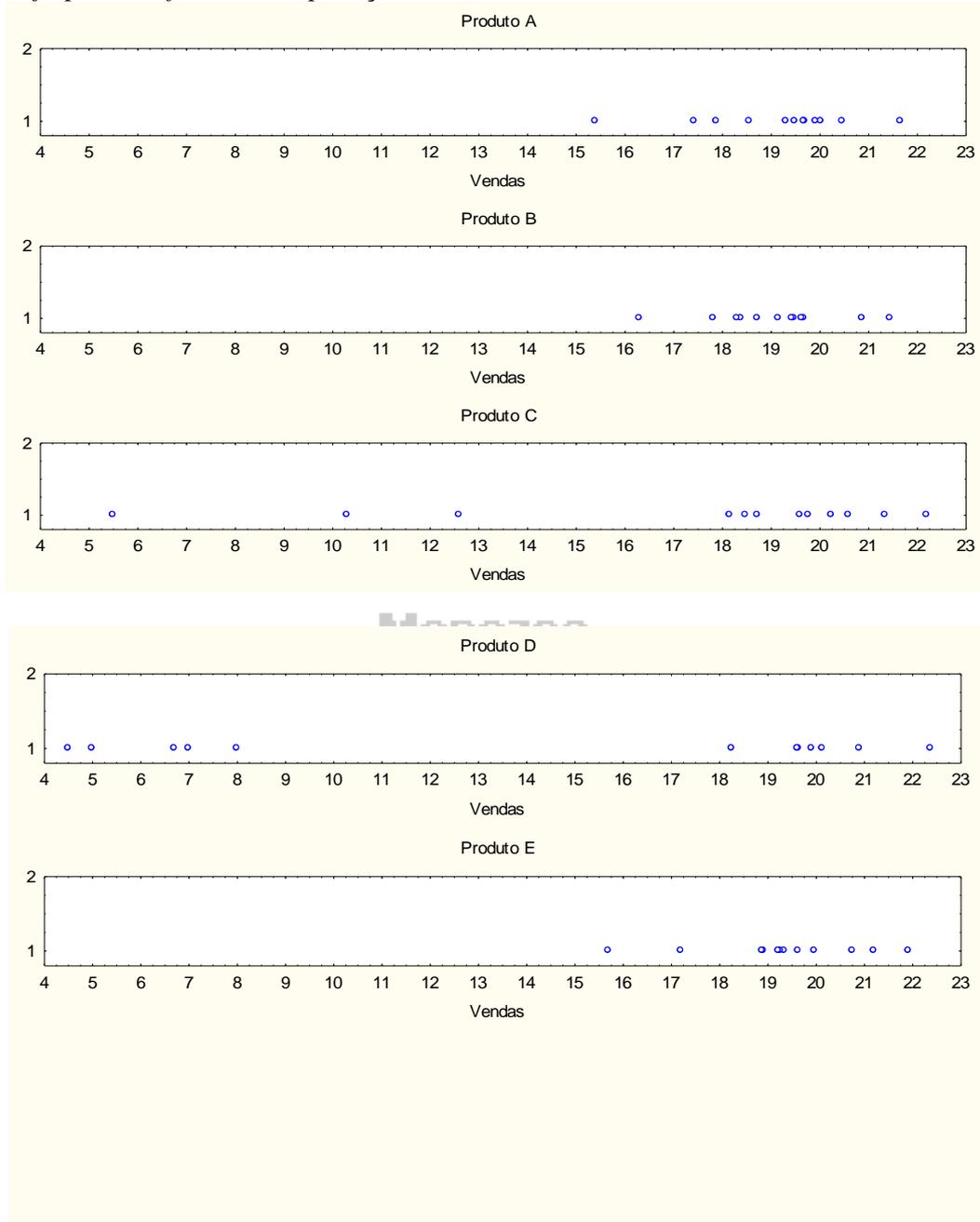
Pelo histograma, ou pelo diagrama de pontos as conclusões são semelhantes às da tabela de frequências agrupadas em classes. A maioria dos percentuais de crescimento está abaixo de 10%, portanto a promoção não teve o efeito esperado.

d) Medidas de síntese. Como há valores discrepantes, tanto superiores quanto inferiores, que podem vir a distorcer a média dos percentuais opta-se pela mediana.

e) A mediana vale 4,405. O valor típico dos percentuais é 4,4%, substancialmente menor do que o 10% considerado mínimo aceitável, portanto conclui-se que a promoção não teve o efeito esperado.

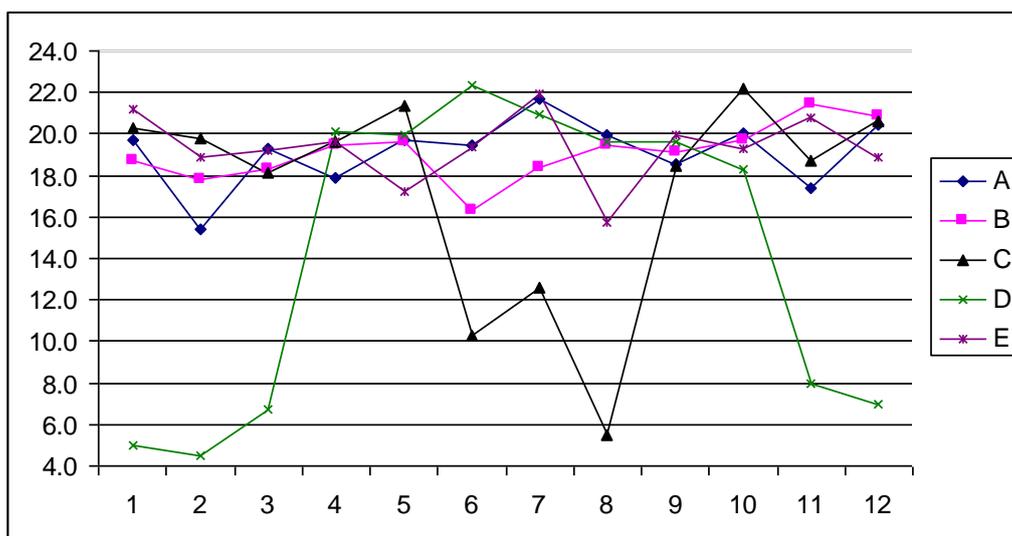
11) O problema da escolha dos 2 produtos que precisam ser retirados de produção.

a) Pode ser usado o diagrama de pontos, porque a variável é quantitativa (com grande variação, podendo ser considerada contínua), e o conjunto de dados é pequeno, apenas 12 observações. É importante ressaltar que é preciso construir um diagrama para cada produto, com a mesma escala, para que seja possível fazer a comparação.



Pelos diagramas de pontos podemos observar que os produtos C e D têm demandas substancialmente diferentes ao longo do ano: em alguns meses são bastante próximas das dos outros produtos A, B e E, mas em outros elas caem muito. Um destes precisa ser escolhido: o produto D tem as vendas menos homogêneas, há 5 valores substancialmente mais baixos, enquanto o produto C tem apenas 3. Como se buscam os produtos com demanda mais homogênea, devemos recomendar a retirada do produto D.

Outra forma de resolver seria utilizando um gráfico de linhas, como o Excel permite fazer.



Os produtos C e D são os que têm os valores mais baixos de vendas, com grande flutuação ao longo dos meses. O produto D permanece por mais tempo nos valores mais baixos de demanda (5 meses), apresentando a demanda menos homogênea. Deve ter, então, sua retirada recomendada.

12) O problema da pesquisa de opinião eleitoral.

a) O candidato Virgulino, porque tem a maior preferência, 36% do total, quase 15% acima do segundo colocado, Elesbão.

b) Mobilidade urbana, pois representa 38% do total, 12% acima do segundo colocado, Honestidade.

c) Sim. Na região SO 43,9% dos eleitores declaram intenção de voto em Virgulino, mesmo percentual encontrado na região NO. Além disso são as regiões com os maiores números de eleitores (244 e 171, respectivamente), representando 40,9% e 28,69% do total, respectivamente.

d) Sim. Se não houvesse, os percentuais por região deveriam ser semelhantes aos encontrados na coluna total: 14,9% para Filisbino, 17,6% para Hermengarda, 21,6% para Elesbão, 35,9% para Virgulino e 9,9% de indecisos. Mas os percentuais são muito diferentes entre as regiões: Filisbino tem 39,2% de preferência na região SE, mas apenas 6,1% na SO; Hermengarda tem 32,9% de preferência na SE, mas apenas 11,9% na SO; Elesbão tem 26,3% de preferência na NO e apenas 16,5% na SE; Virgulino tem 43,9% nas regiões SO e NO, e apenas 11,4% na SE; e o percentual de indecisos varia de 0% (na SE) a 18,9% (na SO).

e) Os que declaram indecisos, pois sua média vale 3,99 salários mínimos, e sua mediana 3,58 salários mínimos, quase quatro vezes maiores do que as medidas de renda dos eleitores do candidato Filisbino.

f) O ideal é calcular os coeficientes de variação percentual, que permite comparar a dispersão de grupos com médias e desvios padrões diferentes.

Candidato	Elesbão	Filisbino	Hermengarda	Virgulino	Indeciso
CV%	27,01%	32,36%	17,18%	47,82%	41,52%

Os eleitores do candidato Virgulino apresentam rendas com maior variabilidade, pois têm o maior coeficiente de variação percentual: 47,82% (desvio padrão representa 47,82% da média).

g) Sim, isso pode ser visto de três formas:

1) Pelos diagramas de pontos, as distribuições das rendas dos eleitores de todos os candidatos (exceto os que estão indecisos) têm claramente um formato assimétrico à direita.

2) A média e a mediana de renda são diferentes entre si em todos os candidatos.

3) Calculando as diferenças  $Quartil\ superior - Mediana$  e  $Mediana - Quartil\ inferior$ , verifica-se que elas são diferentes entre si em todos os candidatos.

13) Estão disponíveis apenas as informações referentes à média e desvio padrão dos salários nas duas faixas etárias.

- a) Com base apenas na tabela a faixa etária "50 ou mais" tem os maiores salários, pois sua média é 398.610 dólares, acima daquela apresentada pela outra faixa etária.
- b) Para avaliar a homogeneidade precisamos calcular os coeficientes de variação dos salários em cada faixa etária, pois tanto médias quanto desvios padrões são diferentes, impossibilitando a comparação direta:

$$cv\%_{\text{Menos de 50}} = \frac{s}{\bar{x}} \times 100 = \frac{219,98}{382,2} \times 100 = 55,56\% \quad cv\%_{50 \text{ ou mais}} = \frac{s}{\bar{x}} \times 100 = \frac{197,92}{398,61} \times 100 = 49,65\%$$

Como o coeficiente de variação da faixa etária "50 ou mais" é MENOR do que o da outra, os seus salários são mais homogêneos (há menor dispersão).

14) A análise dos diagramas em caixa:

a) Valor típico: "50 anos ou mais" apresenta maior valor típico, mediana pouco acima de 350 mil, enquanto o outro grupo apresenta mediana pouco acima de 300 mil.

Assimetria: ambas as distribuições apresentam assimetria, pois as alturas das caixas são diferentes ( $Q_s - Md \neq Md - Q_i$ ), mas o grupo "menos de 50 anos" é mais assimétrico pois a diferença é maior.

Dispersão: o grupo "menos de 50 anos" apresenta maior dispersão, pois a diferença entre os quartis superior e inferior é maior do que no outro grupo.

Valores discrepantes: ambos os grupos possuem valores discrepantes, um no "menos de 50 anos" e dois no outro.

b) Para avaliar qual das faixas de idade devemos levar em conta as três separatrizes principais: quartil inferior, mediana e quartil superior. O grupo "50 ou mais" tem mediana e quartil inferior maiores do que o outro, mas seu quartil superior é menor. Não obstante é possível tomar a decisão apenas pela mediana, indicando que 50% do grupo "50 ou mais" ganham mais do que 350 mil, enquanto que no outro grupo a mediana vale em torno de 300 mil: logo, o grupo "50 ou mais" está sendo melhor remunerado.

15) Crescimento demográfico

Valores típicos: Oriente Médio e África têm os maiores valores típicos, medianas de cerca de 3,0% ao ano. E os menores estão na Europa Oriental e OECD, próximos de zero.

Assimetria: os conjuntos de África e Europa Oriental poderiam ser considerados simétricos, América Latina, OECD e Pacífico/Ásia ligeiramente assimétricos, e o Oriente Médio é assimétrico.

Dispersão: o conjunto com maior dispersão é o Oriente Médio, e os menos dispersos são a Europa Oriental e OECD (demonstrando uma certa homogeneidade demográfica nestas duas regiões).

Valores discrepantes: África e América Latina têm discrepantes inferiores, OECD tem um superior, e as demais regiões não apresentam valores discrepantes.

Média de calorias

Valores típicos: Europa Oriental e OECD têm os maiores valores, na faixa de 3500 calorias diárias, enquanto que a África tem o menor valor, por volta de 2200.

Assimetria: todos os conjuntos são assimétricos, mas Oriente Médio, Pacífico/Ásia e Europa Oriental (onde  $Q_s = Md$ ) são mais do que os outros, a África tem a menor assimetria.

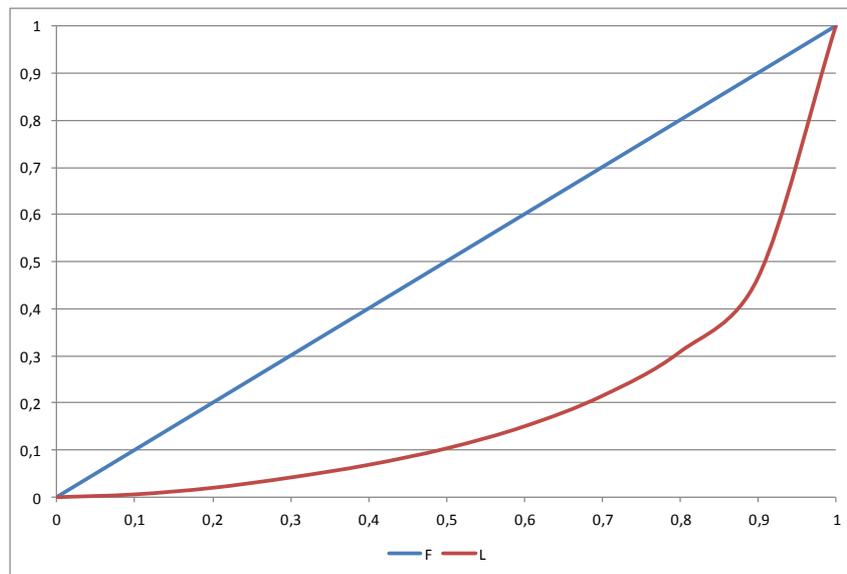
Dispersão: Europa Oriental apresenta a menor dispersão ("caixas" menores), enquanto Pacífico/Ásia apresenta a maior.

É interessante observar o contraste entre os dois diagramas: a África tem um dos maiores valores típicos de crescimento demográfico, e o menor valor típico de calorias ingeridas (indicando um cenário de miséria e fome), enquanto a Europa Oriental e a OECD têm uma situação inversa (o que indica condições socioeconômicas mais favoráveis). Impressiona também a alta taxa de crescimento demográfico no Oriente Médio.

16) Usando o procedimento visto em aula:

Renda	n	F	S	L
-	-	0	-	0
0,6	1	$1/10 = 0,1$	0,6	$0,6/100,6 = 0,005964$
1,4	2	$2/10 = 0,2$	$1,4 + 0,6 = 2$	$2/100,6 = 0,019881$
2,2	3	$3/10 = 0,3$	$2,2 + 2 = 4,2$	$4,2/100,6 = 0,04175$
2,7	4	$4/10 = 0,4$	$2,7 + 4,2 = 6,9$	$6,9/100,6 = 0,068588$
3,5	5	$5/10 = 0,5$	$3,5 + 6,9 = 10,4$	$10,4/100,6 = 0,10338$
4,7	6	$6/10 = 0,6$	$4,7 + 10,4 = 15,1$	$15,1/100,6 = 0,150099$
6,5	7	$7/10 = 0,7$	$6,5 + 15,1 = 21,6$	$21,6/100,6 = 0,214712$
9,4	8	$8/10 = 0,8$	$9,4 + 21,6 = 31$	$31/100,6 = 0,308151$
15,8	9	$9/10 = 0,9$	$15,8 + 31 = 46,8$	$46,8/100,6 = 0,465209$
53,8	10	$10/10 = 1$	$53,8 + 46,8 = 100,6$	$100,6/100,6 = 1$

Plotando o gráfico:



Parece haver concentração de Renda pois a curva de L (a renda analisada) está bem distante de F (igualdade), a área entre as duas, A será grande, fazendo com que o índice de Gini ( $A/A+B$ ) aproxime-se de 1, o que indica maior desigualdade de renda.

17) Com base nas curvas houve redução na desigualdade de renda de 1989 a 1999, pois a curva de 1999 está mais próxima da igualdade do que a de 1989, indicando que a concentração de renda reduziu-se no período.

18) Há indício de redução da desigualdade no período após 1998, pois os índices reduzem-se sistematicamente de 1998 (quando valia 0,567) a 2007 (quando valia 0,52), aproximando-se de zero, que significa igualdade completa de renda.

19)

a) O índice 0,701 corresponde à curva da Renda2, pois ela está mais distante da igualdade, portanto a área entre ela e a curva da igualdade, A, será maior, fazendo com que o índice de Gini ( $A/A+B$ ) aproxime-se mais de 1, o que indica maior desigualdade de renda. Já índice 0,377 corresponde à curva da Renda1, pois ela está mais próxima da igualdade, portanto a área entre ela e a curva da igualdade, A, será menor, fazendo com que o índice de Gini ( $A/A+B$ ) aproxime-se mais de zero, o que indica maior igualdade de renda.

b) Ocorre maior desigualdade de renda no país com Renda2, pois a sua curva está mais distante da igualdade, e o índice de Gini (0,701) está mais próximo de 1.