ESTATÍSTICA DESCRITIVA BÁSICA

- **Para cursos de Pedagogia e Formação de Professores**

Apostila 1
UNIVERSIDADE SEVERINO SOMBRA

Prof. Ilydio Pereira de Sá

ÍNDICE

1) Introdução	**3**
2) Conceitos Iniciais: População, Amostra, Censo, Variáveis	**4**
Exercícios (lista 1)	**8**
Gabarito	**9**
3) Arredondamento e ajuste de dados em Estatística	**10**
4) Noções de amostragem	**11**
Exercícios (lista 2)	**14**
Gabarito	**16**
5) Tabulação e Distribuição de Freqüências	**17**
Exercícios (lista 3)	**30**
Gabarito	**35**

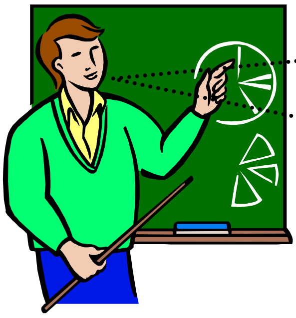
"Nunca se afaste de seus sonhos,
porque se eles se forem
você continuará vivendo,
mas terá deixado de existir."
(Mark Twain)

UNIDADE I – CONCEITOS INICIAIS E TABULAÇÃO

População; Censo; Amostra; Experimento aleatório; Variáveis e atributos; Variáveis aleatórias discretas e contínuas; Normas para apresentação tabular de dados. Organização de Dados Estatísticos. Quadros e tabelas; Distribuição de freqüências; Intervalos de classe; Ponto médio; Freqüências absolutas e relativas; Freqüências acumuladas

1) Introdução:

A Estatística, que é um ramo da Matemática Aplicada, teve suas origens na Antiguidade, relacionando-se basicamente, aos registros e negócios do Estado (daí a origem do nome ***Estatís***tica). Diversos povos tinham o hábito de registrar a população, os nascimentos, os óbitos, as riquezas pessoais, as terras, os impostos, etc.
Na Idade Média colhiam-se informações, geralmente com finalidades tributárias ou bélicas. A partir do século XVI , de uma maneira mais sistemática, começaram a surgir as primeiras tabelas e os primeiros números relativos, associados a fatos sociais.



A palavra **ESTATÍSTICA** vem do termo latino "**STATUS**". Sob essa palavra acumularam-se descrições e dados relativos ao Estado. A ESTATÍSTICA, nas mãos dos estadistas, constitui-se uma verdadeira ferramenta administrativa.

No século XVIII o estudo foi adquirindo feições verdadeiramente científicas e Godofredo Achenwall definiu efetivamente seus objetivos relacionando a Estatística com as demais ciências.

As tabelas tornaram-se mais complexas, surgiram as representações gráficas e o cálculo de probabilidades, e a Estatística deixou de ser uma simples catalogação de dados numéricos coletivos para se tornar o estudo de como chegar a conclusões sobre o todo (População), partindo de observações de partes desse todo (Amostras).

"Feliz aquele que transfere o que sabe e aprende o que ensina."
(Cora Coralina)

2) Conceitos Iniciais: Definições, População, Amostra, Censo, Variáveis.

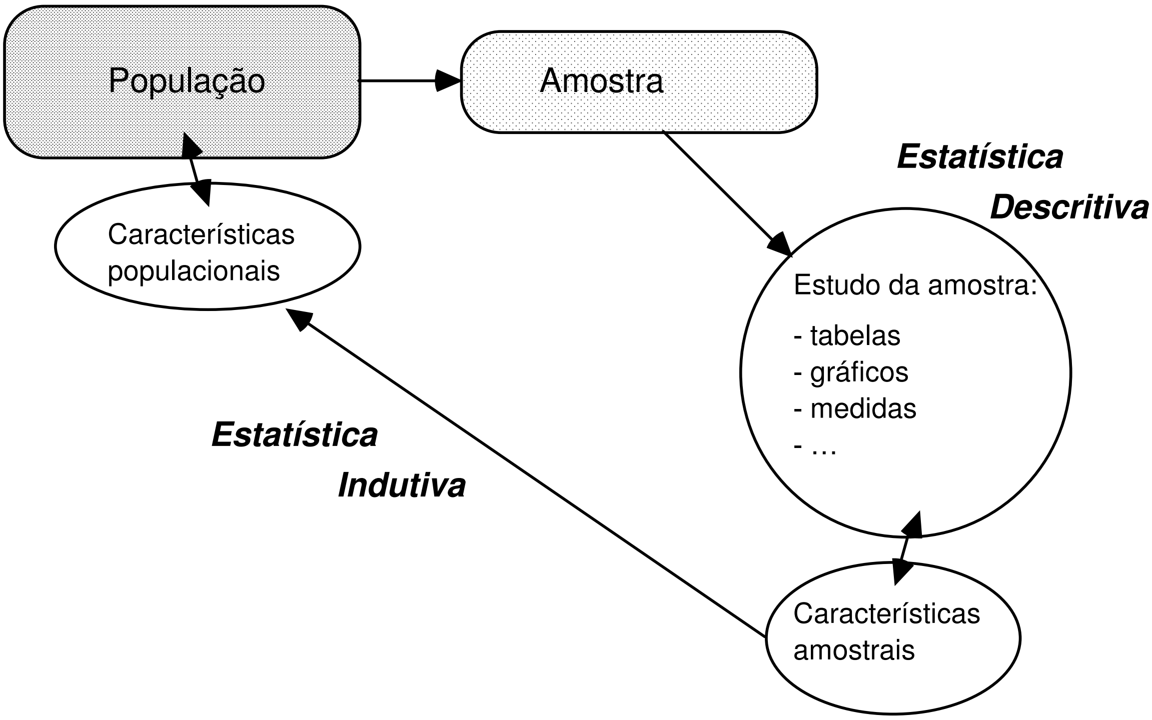
a) **Estatística -"** É um conjunto de métodos e processos quantitativos que servem para estudar e medir os fenômenos coletivos."

"A **Estatística** é uma parte da Matemática Aplicada que fornece métodos para a coleta, organização, descrição, análise e interpretação de dados e para a utilização dos mesmos na tomada de decisões."

b) **Estatística Descritiva X Estatística Indutiva:**

A Estatística Descritiva ou Dedutiva é o ramo da Estatística que tem por objetivo descrever e analisar fatos relacionados a determinado grupo ou população, sem pretender tirar conclusões de caráter mais genérico.

A Estatística Indutiva ou Inferência Estatística , baseando-se nos resultados obtidos da análise de amostra de uma população, procura inferir ou estimar as leis de comportamento de toda a população.



c) **Fases do Método Estatístico:**

Metodologia, normalmente usada na Estatística descritiva, que visa estruturar e organizar as fases que devem ser estabelecidas num estudo estatístico qualquer. As principais fases do método estatístico são: **Coleta de Dados, Crítica dos Dados, Apuração dos Dados, Apresentação dos Dados e Análise dos resultados.**

d) **População e Amostra:**

População é um conjunto de elementos portadores de, pelo menos, uma característica comum.

Devemos tomar cuidado, pois o termo População, em estatística não tem o mesmo significado usual, relacionado apenas a pessoas. Uma população estatística é um conjunto de informações que tenham entre si alguma característica comum. Por exemplo, o conjunto de toas as cores de olhos constitui a população de cores de olhos.
O que importa, nesses casos, é a VARIÁVEL estudada. Por exemplo, uma variável estatística pode ser a RELIGIÃO das pessoas que responderam a uma pesquisa.

Amostra é qualquer subconjunto finito, representativo de uma população. Usamos as técnicas de amostragem quando for muito trabalhoso ou mesmo impossível o estudo completo da população.

Uma amostra, para ser boa, tem de ser representativa, ou seja, deve conter proporcionalmente uma imagem qualitativa e quantitativa o que a população possui.

A amostra deve ser imparcial, isto é, todos os elementos da população devem ter as mesmas chances de fazerem parte da amostra.

Constituída uma amostra, os elementos que a compõem passam a ser tratados como dados estatísticos e podem dar origem as diversas relações ou medidas estatísticas, como, por exemplo: média aritmética, mediana, moda, variância, desvio padrão, coeficiente de variação, etc.

As relações estatísticas possibilitam DESCREVER, sob diversos ângulos, o conjunto de dados representado pela amostra. Por isso, o estudo dessas relações pertence ao campo da **ESTATÍSTICA DESCRITIVA.**

Costumamos, para facilitar, adotar a seguinte convenção de nomenclatura: sempre que fizermos algum cálculo estatístico tendo por base os dados de uma amostra, teremos as chamadas **ESTATÍSTICAS**; sempre que essas relações se referirem à população, teremos os chamados **PARÂMETROS**.

Adotaremos ainda a convenção de usar um **n** (minúsculo) para o tamanho da amostra e um **N** (maiúsculo) para o tamanho da população. Para as relações ou conceitos estatísticos, relacionados às amostras, é usual usarmos letras do nosso alfabeto, enquanto que para os resultados provenientes de toda a população, como nos censos, usamos letras do alfabeto grego.

Vamos tomar como exemplo o importante conceito de DESVIO PADRÃO. Se ele for calculado a partir dos dados da amostra, usaremos o símbolo **s (x),** entretanto, se for calculado a partir dos dados de toda a população, teremos o símbolo **σ (x).**

Por exemplo, num bairro com 5000 habitantes, poderíamos compor uma amostra com 500 pessoas, nesse caso, pela convenção adotada, teríamos:

N = 5000
n = 500
N seria um parâmetro e n seria uma estatística.

É claro que trabalhar com amostras pode ser bastante vantajoso, mesmo que o resultado apresente alguma pequena margem de erro. Por exemplo, sem usar amostras, como poderíamos saber a taxa de defeitos de fabricação de uma fábrica de fósforos, será que seria riscando todos os que foram produzidos?

Para a maioria das pessoas a palavra ***censo*** ou **recenseamento** encontra-se associada à enumeração dos elementos da população de um País. O recenseamento geral de uma população é uma prática que remonta à antiga Roma e Egito, onde temos conhecimento de recenseamentos da população, feitos a intervalos regulares, com o objetivo principal de obter informação para a coleta de impostos, chamada para o serviço militar e outros assuntos governamentais. Apesar disso, a sua prática corrente, com caráter periódico, só teve lugar, na maioria dos Países, a partir do séc. XIX. Esses censos periódicos são feitos em geral de 10 em 10 anos e, em princípio, todos os Países são encorajados a cumprir certas normas internacionais ao elaborar um censo.

PARA VOCÊ RESPONDER ...

Classifique cada um dos exemplos abaixo como sendo casos de Estatística Descritiva ou de Inferência Estatística. Comente brevemente cada caso.

1. Um lote de 100 aparelhos de televisão considera-se em bom estado para venda se, ao serem testados 10 deles, eles não apresentarem qualquer defeito.
2. Uma pesquisa de opinião revelou que 65% da população brasileira apoiava um determinado candidato para Presidente da República. Se esse candidato se apresentar às eleições, é de esperar que ele ganhe.
3. Os 120 empregados de um fábrica ganha em média 400 reais por mês.
4. Baseados numa amostra de 500 trabalhadores de uma empresa de construção civil, acredita-se que a média dos salários dos trabalhadores desse ramo é de 400 reais.
5. Verificadas todas as 5000 peças fabricadas em um dia, por uma Indústria, verificou-se que 4% apresentavam defeitos.

"É possível mentir usando estatísticas, mas se mente mais, e melhor, sem estatísticas. É preciso entender que as amostras podem levar a conclusões erradas. Contudo, as opiniões pessoais, sem base em dados, levam, em geral, a conclusões muito mais erradas."

(*Frederick Mosteller* - Professor em Harvard)

e) **Variáveis:**

Variável é, convencionalmente, o conjunto de resultados possíveis de um fenômeno. A variável pode ser:

Qualitativa ou Atributo (Nominal) ex: sexo, raça, religião, escolaridade, etc. ou

Quantitativa, quando seus valores podem ser expressos por dados numéricos (medidas ou contagem).

A variável quantitativa poderá ser **contínua**, se os seus valores puderem assumir quaisquer resultados dentro de um intervalo qualquer (normalmente decorrente de medidas) ou poderá ser **discreta,** se os seus valores pertencerem a um conjunto enumerável (normalmente decorrente de contagens).

Por exemplo, a quantidade de alunos de um colégio constitui variável discreta, enquanto que o peso desses alunos constitui variável contínua. De um modo geral as contagens dão origem a **variáveis discretas**, enquanto que as medições dão origem a **variáveis contínuas**.

Variável aleatória. Uma variável é denominada **aleatória** quando ela só assume valores decorrentes de experimentos aleatórios. Um experimento é dito aleatório quando, mesmo repetido nas mesmas condições iniciais, pode gerar resultados diferentes e, se o experimento for uniforme todos os valores distintos da variável aparecem em quantidades iguais.

Por exemplo a variável que representa os pontos obtidos no lançamento de um dado (honesto) é aleatória pois os pontos obtidos após cada lançamento poderão ser diferentes, mesmo repetindo-se indefinidamente o experimento de lançar o dado.

Matematicamente, podemos conceituar **variável aleatória** de uma forma mais precisa:

- <u>Variável Aleatória</u>: É o resultado numérico da observação de um fenômeno influenciado por um determinado mecanismo probabilístico. Trata-se da quantificação do resultado produzido por um fenômeno incerto.

Podemos ainda dizer que uma variável aleatória é uma função que associa cada elemento da amostra a um número real qualquer. Por exemplo, ao atirar os dados e registrar os resultados estaremos produzindo uma variável aleatória com extensão total { 1, 2, 3, 4, 5, 6 }. Outro exemplo seria a experiência de escolher uma pessoa ao acaso entre os passantes e medir a sua altura.

É no campo da Estatística Indutiva que o conceito de variável aleatória tem a sua maior aplicação.

Exercícios (Lista 1)

1) Das variáveis abaixo, indique a discreta:
a) O número de "caras" que se pode obter jogando ao ar dez moedas.
b) O tempo de duração de um disco, tomando como unidade o minuto.
c) A temperatura de uma sala, medida em graus Celsius.
d) As alturas dos alunos de uma turma, expressas em cm.
e) As notas dos alunos, em um teste de Matemática.

2) As variáveis discretas são de natureza:
a) quantitativa b) qualitativa c) fracionária d) contínua e) racional

3) As fases básicas do método Estatístico são:
a) Coleta, Crítica, Representação e Análise dos dados.
b) Censo, Planejamento, Representação dos dados.
c) Coleta, Crítica, Apuração, Apresentação dos dados e Análise dos resultados.
d) Planejamento, Divulgação e Análise dos dados.
e) Coleta, Apuração Crítica e Análise dos resultados.

4) Numere a segunda coluna, de acordo com a primeira, e registre a opção correta:
1) Estudo de números associados a fenômenos.
2) Parte da população observada.
3) Denominação dada a atributos ou a quantidades, que variam quanto à grandeza.
4) Grupo de indivíduos ou coisas cujas características são estudadas em forma de um todo, não interessando um elemento em particular.
5) Cada valor observado de uma variável.

( ) Amostra	a) 5 -1 -4 -3 -2
( ) Estatística	b) 2 -3 -4 -1 -5
( ) População	c) 3 -1 -4 -2 -5
( ) Variável	d) 2 -1 -4 -5 -3
( ) Dado	e) 2 -1 -4 -3 -5

5) População ou Universo é:
a) Conjunto de pessoas
b) Conjunto de indivíduos que apresentam características especiais.
c) Conjunto de elementos que apresentam uma característica comum.
d) Subconjunto confiável para um estudo qualquer.
e) Nada disso.

6) O método aplicado por Institutos de Pesquisa, nas prévias eleitorais, pertence ao ramo da:
a) Estatística Descritiva
b) Estatística Indutiva
c) Estatística Aplicada
d) Estatística Geral
e) Estatística Dedutiva.

7) Ao nascer, os bebês são pesados e medidos, para saber se estão dentro das tabelas padrão. Estas duas variáveis são:
a) qualitativas
b) discretas
c) contínuas
d) contínua e discreta, respectivamente

8) Parcela da população convenientemente escolhida para representá-la:
a) variável b) rol c) dados d) amostra e) atributo

9) É exemplo de atributo:
a) número de filhos b) estado civil c) altura d) peso e) idade

10) É exemplo de variável discreta:
a) Número médio de filhos, por família de uma localidade.
b) Salário de uma pessoa em dólares.
c) Altura média das montanhas de uma cidade.
d) Votos anulados em uma seção eleitoral.
e) Porcentagem de acertos ao alvo, de um atirador.

11) Qual dos exemplos a seguir **não apresenta** um caso de variável aleatória?
a) número de gols feitos por um artilheiro, após cada partida de um campeonato de futebol.
b) nota de um aluno nas prova de matemática que irá realizar ao longo de um ano letivo.
c) taxa de inflação mensal brasileira.
d) número de dias do mês de março, ao longo de uma determinada década.
e) número de dias chuvosos ao longo do mês de janeiro de 2006.

GABARITO

01) A 02) A 03) C 04) E
05) C 06) B 07) C 08) D
09) B 10) D 11) E

3) Arredondamento e Ajuste de Dados em Estatística:

Quando for necessário, em cálculos e tabelas estatísticas, o arredondamento de dados, devemos proceder de acordo com a resolução número 886/66, da Fundação IBGE.

I) Quando o primeiro algarismo a ser abandonado for 0, 1,2,3 ou 4, fica inalterado o último algarismo a permanecer.

exemplos:
a) 53,24 arredondado ao décimo mais próximo será 53,2.
b) 12,473 arredondado ao centésimo mais próximo será 12,47.

II) Quando o primeiro algarismo a ser abandonado for 6,7,8 ou 9, aumenta-se de uma unidade o último algarismo a permanecer.

exemplos:
a) 42,87 arredondado ao décimo mais próximo será 42,9.
b) 13,7 arredondado ao inteiro mais próximo será 14.

III) Quando o primeiro algarismo a ser abandonado for o 5, há duas situações a considerar:

A) Se ao 5 seguir, em qualquer ordem, um algarismo diferente de zero, aumenta-se de uma unidade o algarismo a permanecer.

exemplos:
a) 2,352 arredondado a décimos será 2,4
b) 14,325004 arredondado a centésimos será 14,33.

B) Se o 5 for o último algarismo , ou ao 5 só se seguirem zeros, o último algarismo a ser conservado só será aumentado de uma unidade se for ímpar.
exemplos:
a) 24,75 arredondado a décimos será 24,8.
b) 14,7850 arredondado a centésimos será 14,78.

Compensação ou Ajuste de Dados:

Quando houver arredondamento em somatórios e precisarmos "ajustar" alguns elementos, de modo a acertar a soma, usamos "descarregar" a diferença na maior ou nas maiores parcelas.

exemplo:
25,32 + 17,85 + 10,44 + 31,17 = 84,78.

Se arredondássemos todas as parcelas e a soma, para décimos, teríamos:
25,3 + 17,8 + 10,4 + 31,2 = 84,8.

No entanto esta soma agora é 84,7 que não é o arredondamento correto de 84,78. A compensação faríamos na maior parcela (31,2), tornando-a 31,3, logo, teríamos:
25,3 + 17,8 + 10,4 + 31,3 = 84,8.

4) Noções de Amostragem:

Para a escolha de uma amostra representativa de uma população, devemos utilizar técnicas especiais, de modo a garantir, tanto quanto possível , o acaso na escolha.
Sendo assim, cada elemento da população passa a ter a mesma chance de ser escolhido, o que garante à amostra o caráter de representatividade.

As principais técnicas de amostragem são:

A) Amostragem Casual, Aleatória Simples ou Randômica:

Este tipo de amostragem é equivalente a um sorteio lotérico. Na prática, a amostragem casual pode ser realizada numerando-se a população de 1 a n e sorteando-se por um método confiável qualquer os k elementos que constituirão a amostra.
Quando o número de elementos da amostra for muito grande, esse tipo de sorteio se torna muito trabalhoso. A fim de facilitá-lo foi elaborada uma tabela - **A tabela de números aleatórios -** construída de modo que os 10 algarismos do nosso sistema de numeração são distribuídos ao acaso nos números formadores das linhas e das colunas.
Obs: Modernamente, os computadores possuem programas de seleção aleatória ou randômica, que substituem com vantagem o uso da tabela.

B) Amostra Proporcional Estratificada:

Muitas vezes a população divide-se em subpopulações, - **os estratos-**. Na técnica de amostragem proporcional, considera-se a existência de tais estratos e obtém-se os elementos da amostra como uma divisão proporcional direta ao número de elementos de tais estratos.

Exemplo: Como ficaria a divisão da amostra de 10% de uma população de 90 alunos, com 54 meninos e 36 meninas, por amostragem proporcional estratificada.

Solução:

Como a amostra é de 10 % da população, teremos de calcular 10 % de cada um dos estratos, levando-se em conta os arredondamentos adequados.
Meninos : 10 % de 54 = 5,4 logo 5 meninos.
Meninas : 10 % de 36 = 3,6 logo 4 meninas.

C) Amostragem Sistemática:

Quando os elementos da população já se acham ordenados, não há necessidade de construir o sistema de referência. São exemplos os fichários de um médico, os prédios de uma rua, as linhas de produção, etc. Nestes casos , a seleção dos elementos que constituirão a amostra pode ser feita por um sistema imposto pelo pesquisador. A esse tipo de amostragem denominamos **sistemática.**

O método mais usado na amostragem sistemática é a formação de progressões aritméticas de razão k, sendo k = N : A, onde N é o número de elementos da população e A é a quantidade de elementos da amostra.. O primeiro elemento da P.A fica determinado por sorteio.

Exemplo:

Num consultório médico, os clientes estão cadastrados em fichas numeradas de 1 a 500. Compor uma amostragem sistemática com 10 elementos, começando pela ficha número 12.

Solução:

Razão da P.A: k = 500 : 10 = 50.

Amostra: Clientes números: 12, 62, 112, 162, 212, 262, 312, 362, 412, 462.

Uma questão importante ...

Qual o tamanho ideal que deve ter uma amostra?

Um problema relevante que temos, a partir da escolha de uma amostra é o de saber qual a *dimensão* desejada para a amostra a recolher.

Este é um problema para o qual nesta fase, não é possível avançar nenhuma teoria, já que depende de estudos referentes à Estatística Indutiva.

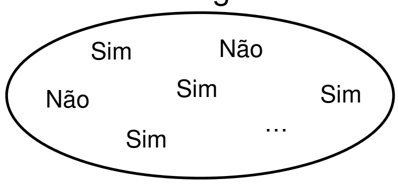
Pode começar por dizer que, para se obter uma amostra que permita calcular estimativas suficientemente precisas dos parâmetros a estudar, a sua dimensão depende muito da variabilidade da população estudada. Por exemplo, se relativamente à população constituída pelos alunos da 7º série de uma escola de ensino fundamental, estivermos interessados em estudar a sua média de idades, a dimensão da amostra a recolher não necessita de ser muito grande já que a variável idade apresenta valores muito parecidos, numa classe etária muito restrita. No entanto se a característica a se estudar fosse o tempo médio que os

alunos levam a chegar de casa à escola, a amostra teria de ter uma dimensão maior, uma vez que a variabilidade da população é muito maior. Cada aluno pode apresentar um valor diferente para esse tempo.
Chamamos a atenção para a existência de técnicas que permitem obter valores mínimos para os tamanhos das amostras a recolher e que garantem estimativas com uma determinada precisão desejada. Uma vez garantida essa precisão, a opção por escolher uma amostra de maior dimensão, é uma questão a ponderar entre os custos envolvidos e o ganho com o acréscimo de precisão.

Convém ainda observar que a dimensão da amostra a recolher não é diretamente proporcional à dimensão da população a estudar. Se,por exemplo para uma população de dimensão 1000 uma amostra de dimensão 100 for suficiente para o estudo de determinada característica, não significa necessariamente uma amostra de dimensão 200 para estudar a mesma característica de uma população análoga, mas de dimensão 2000.

Um exemplo:

O Senhor R, candidato à Câmara da cidade do Rio de Janeiro, pretende saber, qual a percentagem de eleitores que pensam votar nele nas próximas eleições. Havendo algumas limitações de tempo e dinheiro, a empresa encarregada de fazer o estudo pretendido decidiu recolher uma amostra de dimensão 1000, perguntando a cada eleitor se pensava em votar no senhor R. Como resultado da amostragem obteve-se um conjunto de sim's e não's, cujo aspecto não é muito agradável, pois à primeira vista não conseguimos concluir nada:



Procede-se à redução dos dados, resumindo a informação sobre quantos sim's se obtiveram, chegando-se à conclusão que nas 1000 respostas, 635 foram afirmativas. Então dizemos que a percentagem de eleitores que pensam votar no candidato, de entre os consultados, é de 63,5%.

A função da Estatística Descritiva acabou aqui! (Se toda a População tivesse sido inquirida, este estudo descritivo seria suficiente para dar a resposta à consulta feita).

Poderemos agora inferir que 63,5% dos eleitores da cidade do Rio de Janeiro pensam votar no Senhor R?

A resposta a esta pergunta nem é sim, nem não, mas talvez. É agora que temos necessidade de utilizar o conceito de Probabilidade, para quantificar a incerteza associada à inferência. Assim, existem processos de inferência estatística que, do resultado obtido a partir da amostra, nos permitirão concluir que o intervalo [60,5%, 66,5%] contém o valor exato para a percentagem de eleitores da cidade que pensam votar no Senhor R, com uma confiança de 95%.

Nota - A confiança de 95% deve ser entendida no seguinte sentido: se recolhermos 100 amostras, cada uma de dimensão 1000, então poderemos construir 100 intervalos; destes 100 intervalos esperamos que 95 contenham o verdadeiro valor da percentagem (desconhecida) de eleitores da cidade do Rio de Janeiro, que pensam votar no senhor R.

Exercícios (Lista 2)

12) (Fiscal do ICMS - 89)
Em uma amostragem sistemática, de tamanho 50, de uma população de 2000 elementos, o primeiro elemento selecionado é o 16. Os dois elementos seguintes a serem escolhidos são:
a) 32 e 48 b) 50 e 66 c) 50 e 100 d) 56 e 96 56 e 106

13) Numa escola estão matriculados 280 meninos e 320 meninas (não existindo alunos irmãos). O diretor da escola, desejoso de conhecer as condições de vida extra-escolar de seus alunos e não dispondo de tempo para entrevistar todas as famílias, resolveu fazer um levantamento por amostragem proporcional estratificada, composta de 50 alunos. As famílias de quantos meninos serão entrevistadas?
a) 23 b) 42 c) 25 d) 27 e) 28

14) Uma cidade A , apresenta o seguinte quadro, relativo ao número de estudantes de suas escolas de Ensino Fundamental.

Escolas	Masculino	Feminino
X	120	130
Y	240	360
Z	180	420
T	300	200
Total	840	1110

Obtenha uma amostra proporcional estratificada de 120 estudantes e indique quantos meninos representarão a escola Z
a) 15 b) 11 c) 14 d) 20 e) 32

"É impossível a existência de uma sociedade formada por pessoas que não sonham. Elas estariam mortas em duas semanas."

William Burroughs - escritor americano

15) Tendo-se os números: 2,483 ; 48,6 e 368, arredonde-os, utilizando-se dos seguintes critérios, respectivamente:

- **centésimo mais próximo**
- **unidade mais próxima**
- **centena mais próxima**

(a) 2,48 - 48 - 400
(b) 2,49 - 49 - 300
(c) 2,4 - 48 -300
(d) 2,48 - 48 - 300
(e) 2,48 - 49 – 400

16) Arredondando-se para inteiros, e devendo o total ser 100 %, na distribuição percentual imaginária abaixo, qual será o valor correspondente ao estado do Amazonas?

Amazonas	48,4 %
Pará	15,3 %
Maranhão	13,8 %
Piauí	9,1 %
Ceará	13,3 %
Total	99,9 %

a) 48 % b) 47 % c) 49 % d) 50 % e) 45%

17) Entre as afirmativas seguintes, assinale a que é **FALSA.**

(a) Faz-se um levantamento por censo quando todos os elementos da população são pesquisados.
(b) Faz-se um levantamento por amostragem quando se pesquisa parte de uma população e, com base no subconjunto pesquisado, pode-se tirar conclusões sobre toda a população.
(c) A decisão entre os tipos de levantamento a serem realizados, censo ou amostragem, dependerá, entre outras coisas, do prazo de realização da pesquisa e dos recursos financeiros disponíveis.
(d) As afirmativas I, II e III são falsas.

18**) (TCU)** Assinale a opção correta:
(a) Estatística inferencial compreende um conjunto de técnicas destinadas à síntese de dados numéricos.
(b) O processo utilizado para se medir as características de todos os elementos de uma dada população recebe o nome de censo.
(c) A Estatística Descritiva compreende as técnicas por meio das quais são tomadas decisões sobre uma população com base na observação de uma amostra.
(d) Uma população só pode ser caracterizada se forem observados todos os seus componentes.
(e) Parâmetros são medidas características de grupos, determinadas por meio de uma amostra aleatória.

19) Uma amostragem sistemática será constituída de 6 elementos, para uma população de 360 elementos. O primeiro elemento sorteado para a amostra foi o 45. Se todos os elementos do universo são numerados de 1 a 360, qual será o terceiro elemento dessa amostra?
a) 165 b) 105 c) 205 d) 245 e) 145

20) Assinale a alternativa FALSA:
(a) A Estatística é constituída de um conjunto de métodos e processos quantitativos que servem para estudar e medir os fenômenos coletivos.
(b) A população constituída por todos os parafusos produzidos numa fábrica em um certo dia é finita, enquanto que a população constituída de todos os resultados (cara ou coroa) em sucessivos lances de uma moeda é infinita.
(c) Um censo constitui-se pelo exame de todas as unidades de uma população finita.
(d) No estudo de determinada característica associada a uma população, deve-se recorrer a uma amostra quando for impraticável (ou mesmo impossível) observar todo o grupo.
(e) Experimento aleatório é o conjunto de todos os resultados possíveis de uma experiência aleatória.

GABARITO

12) D 13) A 14) B 15) E

16) C 17) D 18) B 19) A 20) E

Dificuldades reais podem ser resolvidas;
apenas as imaginárias são insuperáveis."

Theodore N. Vail

5) Tabulação

Introdução:

Feita a coleta e a crítica dos dados, com a respectiva apuração, devemos efetuar a apresentação desses dados, o que é feito normalmente através de **tabelas ou gráficos.** Trataremos inicialmente das tabelas, e, um postulado básico que se deve obedecer é que: ***"Uma tabela deve propiciar um máximo de esclarecimentos com um mínimo de espaço e tempo."***

Estruturação:

No Brasil as regras para construção de tabelas e gráficos são regidas pelo IBGE. Uma tabela é constituída de título, cabeçalho, corpo, coluna indicadora e rodapé. No cabeçalho devemos ter as informações sobre o que está apresentado na tabela (o que, onde e quando...). No corpo da tabela temos as células, onde estarão os dados representativos do fenômeno estudado. No rodapé, recomenda-se que se coloque a fonte das informações.

Consta ainda da norma específica que as tabelas não sejam fechadas nas extremidades (esquerda e direita).

Exemplo: Tabela 1
Casos registrados de intoxicação humana, segundo a causa dominante. Brasil, 1993.

Título — Cabeçalho — Coluna indicadora — Corpo — Rodapé

Causa	Casos
Acidente	29 601
Abuso	2 604
Suicídio	7 965
Profissional	3 735
Outras	1 959
Ignorada	1 103

Fonte: MS/FIOCRUZ/SINITOX

No corpo de uma tabela , que é representado por uma série de colunas e subcolunas, estão colocados os dados apurados. O corpo pode ser dos tipos: ***Entrada Simples; Dupla Entrada (contingência) e de Múltipla Entrada.***

No rodapé da tabela devemos colocar as legendas e convenções usadas, fornecer a fonte dos dados, de modo a dar maior autenticidade à tabela e, se necessário, as observações que se fizerem necessárias.

"Ter razão é fácil.
Perceber que os outros a têm - eis o problema."
(M. Silva Brito)

Exemplos:
Entrada Simples

Previsão populacional
Cidade de São Paulo: 1984 - 2020.

Anos	População
1984	9 439 000
1990	11 160 000
1995	12 224 000
2000	13 410 000
2010	14 910 000
2020	15 532 000

Fonte: Sabesp

Entrada Dupla

Número de Funcionários
Empresa X - 1984

	Homens	**Mulheres**	**Total**
Maiores	160	130	290
Menores	140	110	250
Total	300	240	540

Observação:
De acordo com a deliberação 886, da fundação IBGE, nas casas ou células de uma tabela valem as convenções:

- Um traço horizontal (-) quando o valor do dado for zero.
- Três pontos (...) quando não temos os dados.
- Um ponto de interrogação(?) quando temos dúvida quanto à exatidão de determinado valor.
- Zero (0) quando o valor é muito pequeno para ser expresso pela unidade utilizada.

Séries Estatísticas:

Denominamos **série estatística** a toda tabela que apresenta a distribuição de um conjunto de dados estatísticos, em função da época, do local ou da espécie dos dados.

Disso, podemos concluir que as variáveis: **tempo, local e espécie** são os componentes fundamentais de tais séries, e elas poderão ser denominadas:**Históricas, Geográficas ou Específicas,** respectivamente, dependento do elemento variável.

- **Série Histórica, Cronológica , Temporal ou Marcha** .. A variável é o tempo.
- **SérieGeográfica, Espacial, Territorial ou de localização. ..**A variável é o local.
- **Série Específica , Qualitativa ou Categórica** .. A variável é o fato ou categoria, permanecendo fixos o local e o tempo.

- **Séries Mistas ou Conjugadas.** São composições de duas ou mais das anteriores.

Exemplo:
Classificar as séries estatísticas, representadas pelas tabelas seguintes:

A)Produção Brasileira /Carvão Mineral Bruto (79/81)

ANO	QUANTIDADE (x 1000 t)
1979	13 943
1980	16 006
1981	17 434

Fonte: Ministério das Minas e Energia.

Resposta: É uma série do tipo..

B)Avicultura no Brasil (1980)

Espécie	Quantidade
Galinhas	447 411 000
Patos	4 887 000
Perus	2 074 000
Codornas	831 000

Fonte: IBGE

Resposta: É uma série do tipo: ....

C) Produção de Celulose(1981)

Estado	Produção (t)
São Paulo	958 569
Espírito Santo	400 760
Paraná	339 569
S. Catarina	323 812
Minas Gerais	226 559

Fonte: Associação Nacional de Fabricação de Papel e Celulose.

Resposta: É uma série do tipo ....

D) Estimativa de Renda Interna Segundo os Ramos de Atividade (1994/96)

Renda Interna (Em milhares de dólares)

Ramos de Atividade	1994	1995	1996
Agricultura	421 933	708 848	1 446 050
Indústria	1 046 289	1 726 161	3 778 060
Serviços	1 662 867	2 886 801	5 880 469
Total	3 131 089	5 321 810	11 104 579

Fonte: Fundação Getúlio Vargas

Resposta: É uma série do tipo
....

Dados Absolutos e Dados Relativos.

Os dados estatísticos resultantes da coleta direta da fonte, sem outra manipulação senão a contagem ou medida, são chamados **dados absolutos.**
Dados relativos são resultantes de razões que se estabelecem entre dados absolutos, e têm por finalidade a comparação entre quantidades.
Os dados relativos mais comuns em estatística são : **porcentagens, índices , coeficientes e taxas.**

A) As Porcentagens:

O emprego de porcentagens é de grande valia, principalmente quando queremos destacar a participação da parte no todo e a consequente comparação de grandezas.
Podemos estabelecer, em tais casos, a fórmula:

$$\boxed{P = \frac{p \times 100}{T}}$$

Onde P é a porcentagem procurada, p é o dado numérico referente à parte e T é o dado numérico referente ao todo (referencial adotado).

Obs: Como é um cálculo que envolve divisão, enfrentaremos os arredondamentos e compensações que já estudamos anteriormente. Normalmente arredondamos a décimos as porcentagens.

Exemplo:

Matrículas nas Escolas da Cidade X em 2002

Categoria	Quantidade de alunos	%
Educação Infantil	34 600	36,9
Ensino Fundamental	45 567	48,6
Ensino Médio	13 560	14,5
Total	93 727	100,0

Fonte: Dados Imaginários

- Educação Infantil = 34 600 X 100 : 93 727 = 36,915 % ~ 36,9 %
- Ensino fundamental = 45 567 X 100 : 93 727 = 48,616 % ~ 48,6 %
- Terceiro grau = 13 560 X 100 : 93 727 = 14,467 % ~ 14,5 %

B) Os índices , os coeficientes e as taxas.

Os **índices** são razões entre duas grandezas tais que uma não inclui a outra.
Exemplo: **Densidade demográfica** = $\frac{\text{População}}{\text{Superfície}}$

Os **coeficientes** são razões entre duas grandezas tais que uma delas está contida na outra, ou seja é uma razão entre parte e todo.
Exemplo: Coef. de natalidade = $\frac{\text{Nº de nascimentos}}{\text{População total}}$

As **taxas** são obtidas multiplicando-se por uma potência de 10 (10, 100, 1000, etc.) os resultados dos coeficientes ou índices.

A mais usada é a taxa percentual ou porcentagem.
Exemplo:
Calcule a taxa de aprovação de uma classe com 40 alunos, dos quais foram aprovados 36.

$$TA = \frac{36}{40} \text{ x } 100 = 0{,}9 \cdot 100 = \mathbf{90\ \%}$$

As Distribuições de Freqüência.

É um tipo de série estatística muito especial, qualitativa ou específica, onde os dados estão agrupados em faixas, de acordo com as suas repetições, podendo-se utilizar faixas ou intervalos que contenham tais valores. O número de observações ou repetições de um determinado dado é a sua **freqüência**.
As distribuições de Freqüência dividem-se em dois tipos:

I) Tipo A ou Discreta.

Normalmente usamos este tipo de distribuição quando não há um número muito grande de valores distintos da variável.
É uma tabela com duas colunas básicas:(x_i), que representa os valores distintos da variável e (f_i), que indica o número de repetições de cada valor (freqüência absoluta).

Exemplo:

Um professor anotou as notas de Matemática de uma classe de 30 alunos, obtendo os seguintes dados brutos (ainda não organizados):

07	**06**	**05**	**08**	**09**	**10**
08	**06**	**08**	**07**	**05**	**06**
10	**09**	**09**	**08**	**07**	**05**
08	**06**	**06**	**07**	**04**	**04**
06	**05**	**10**	**09**	**04**	**05**

Fazendo a distribuição de freqüência discreta (tipo A) para essas notas, teremos:

x_i	f_i
04	3
05	5
06	6
07	4
08	5
09	4
10	3
$\sum$	30

Obs: Usamos também o cálculo da **freqüência relativa (f_r)**, que é a razão entre a freqüência absoluta (f_i) e o número total de elementos observados N ou $\sum$ f, e que pode também ser expressa em forma de porcentagem (taxa relativa)

$$\boxed{f_r = \frac{f_i}{N}}$$

Exemplo:

Distribuição de profissões entre pacientes potencialmente suicidas

Profissão (x_i)	Freqüência (f_i)	Freq. Relativa (fr)	Taxa (%)
Serviços Gerais*	75	0,249	24,9
Doméstica**	55	0,182	18,2
Do Lar	53	0,175	17,5
Indeterminada	29	0,096	9,6
Emprego especializado***	23	0,076	7,6
Menor	20	0,066	6,6
Desempregado	15	0,050	5,0
Estudante	14	0,046	4,6
Lavrador	12	0,040	4,0
Autônomo	4	0,013	1,3
Aposentado	2	0,007	0,7
Total	302	1	100

** garçom, encanador, pedreiro, frentista, operário, padeiro, açougueiro, borracheiro etc.*
*** copeira, faxineira, costureira e bordadeira*
**** enfermeira, modelo, protético, escrivão, professor e vendedor*

A tabela mostra a distribuição de pacientes potencialmente suicidas. A maior incidência foi observada entre profissionais mal remunerados e com sobrecarga de trabalho, sem perspectiva de ascensão social. O alto percentual entre menores e estudantes (6,6% + 4,6%) confirma o fato relatado na literatura de que o índice de tentativas de suicídio entre adolescente é preocupante. Tal tipo de distribuição é o que denominamos **tabela de freqüência discreta.**

II) Tipo B ou Contínua:

Quando o número de valores distintos da variável for muito grande, o processo anterior seria, além de trabalhoso, desvantajoso em termos de apresentação já que exigiria um espaço muito grande. Dessa forma, é usual agruparmos os valores distintos obtidos em intervalos ou classes, de modo a resumir a tabela , em tempo de elaboração e forma de apresentação.
A forma mais comum de representarmos uma classe de valores é **a |—— b,** que significa uma classe com todos os valores entre **a (inclusive)** e **b (exclusive).**

Exemplo:
Os valores abaixo relacionados em rol (estão organizados), representam as alturas dos 40 alunos de uma turma, em cm. Organizar uma distribuição de freqüências do tipo B (Contínua), com 6 classes distintas.

150	**154**	**155**	**157**	**160**	**161**	**162**	**164**	**166**	**169**
151	**155**	**156**	**158**	**160**	**161**	**162**	**164**	**167**	**170**
152	**155**	**156**	**158**	**160**	**161**	**163**	**164**	**168**	**172**
153	**155**	**156**	**160**	**160**	**161**	**163**	**165**	**168**	**173**

Solução:

Estaturas	f_i
150 |—— 154	04
154 |—— 158	09
158 |—— 162	11
162 |—— 166	08
166 |—— 170	05
170 |—— 174	03
$\sum$	40

Importante:
Existem dois princípios básicos que devem nortear a elaboração de uma distribuição contínua de freqüências:

I. **As classes devem ser exaustivas, ou seja, todos es valores da variável devem estar inseridos em uma das classes.**
II. **As classes devem ser mutuamente exclusivas, ou seja, nenhum valor deve pertencer a duas classes ao mesmo tempo.**

Vamos agora completar o nosso estudo de distribuição de freqüências com mais algumas **definições** importantes ao nosso estudo.

(a) **classes** - são os intervalos de variação da variável, no exemplo anterior foram feitas 6 classes.
(b) **limites de classe** - são os extremos de cada classe. O limite inferior representaremos por ℓ e o limite superior, representaremos por **L .**
Por exemplo, na tabela acima, os limites da terceira classe são: ℓ **= 158 e L = 162.**

(c) **amplitude do intervalo de classe, ou intervalo** é o número que expressa, no caso de intervalos fechados em apenas um dos extremos (à esquerda, segundo deliberação do IBGE, para o Brasil), a diferença entre os limites superior e inferior.
No nosso exemplo ainda, cada intervalo tem amplitude igual a 4.

(d) **amplitude total da distribuição** é a diferença entre o limite superior da última classe e o limite inferior da primeira classe, para intervalos semifechados, ou seja:

$$\textbf{ATD = L (máx.) - } \ell \textbf{ (mín.)}$$

Em nosso exemplo, ATD = 174 - 150 = 24 cm.

É imediato observar que, se as classes possuem o mesmo intervalo **h,** com número de classes igual a **k**, vale a relação:

$$\mathbf{h = ATD : k}$$

No exemplo dado , temos: h = 24: 6 = 4.

Observe que a determinação dos parâmetros básicos da distribuição de freqüências depende do número de classes a ser considerado. Apesar de ser uma quantidade à escolha do pesquisador, existem algumas fórmulas empíricas que pretendem "amarrar" o número de classes ao número de elementos observados n. A mais usada é a fórmula de **Sturges,** que é:

FÓRMULA DE STURGES

$$\boxed{\mathbf{k \cong 1 + 3{,}3 \,.\, \log n}}$$

No nosso exemplo, teríamos: $\mathbf{k = 1+3{,}3 \,.\, \log 40 = 1 + 3{,}3 \,.\, 1{,}6 \cong 6{,}28}$ ou **6 classes.**

Nesse tipo de exercício o valor do logaritmo decimal é normalmente dado no problema, como no nosso caso que log 40 ≅ 1,6.

Há quem use também a relação $\mathbf{k} \cong \sqrt{n}$, que, ainda no nosso exemplo geraria o resultado: $k \cong \sqrt{40} \cong 6$ classes.**(Procedimento da raíz)**

(e) **Amplitude amostral** é a diferença entre o valor máximo e o valor mínimo da amostra de valores ou população observada. No exemplo a amplitude amostral é 23, pois esta é a diferença entre o valor máximo 173 e o valor mínimo 150. Note que na elaboração da tabela, aumentamos a amplitude para 24, para que a divisão por 6 fosse exata e facilitasse a montagem.

(f) Ponto médio da classe é o valor que normalmente representa a classe no uso de fórmulas estatísticas, e é o ponto que divide o intervalo em duas partes iguais. A obtenção do ponto médio pode ser feita por: $\mathbf{PM = (\ell + L) : 2}$ **ou** $\mathbf{\ell + h : 2.}$

(g) **Freqüência acumulada** pode ser crescente, quando representa a soma das freqüências de uma classe com as anteriores (também chamada de freqüência “abaixo de”) ou decrescente, quando é a soma das freqüências de uma classe com as posteriores (também chamada de freqüência “acima de”). Este conceito representa a quantidade de elementos que antecedem ou que sucedem um certo dado é terá grande uso posteriormente, no capítulo das medidas de posição em estatística.

Voltando ao nosso exemplo, vamos agora completar a tabela com os demais conceitos que estudamos:

	f_i	$f_i r$	$F_i r$ %	F_{ac} cr	F_{ac} dc
150 ├── 154	**04**	**0,10**	**10**	**04**	**40**
154 ├── 158	**09**	**0,22**	**22**	**13**	**36**
158 ├── 162	**11**	**0,28**	**28**	**24**	**27**
162 ├── 166	**08**	**0,20**	**20**	**32**	**16**
166 ├── 170	**05**	**0,12**	**12**	**37**	**8**
170 ├── 174	**03**	**0,08**	**08**	**40**	**3**
$\sum$	**40**	**1,00**	**100**		

Outro exemplo: Distribuição do Nível de colesterol (mg/dl) em 80 indivíduos.

	Freqüência Absoluta (f_i)		Freqüência Relativa	
Nível (classes)	Simples	Acumulada	Simples	Acumulada
100 |── 150	2	2	0,0256	0,0256
150 |── 200	24	26	0,3077	0,3333
200 |── 250	35	61	0,4487	0,7821
250 |── 300	14	75	0,1795	0,9615
300 |── 350	1	76	0,0128	0,9744
350 |── 400	1	77	0,0128	0,9872
400 |── 450	0	77	0,0000	0,9872
400 |── 500	1	78	0,0128	1,0000
	78		1	

Resumindo as etapas para a elaboração de uma distribuição de freqüências, do tipo contínua:

1) Encontrar o máximo (maior) e o mínimo (menor) valores do conjunto de dados; obter a diferença entre esses valores extremos, que chamaremos de amplitude total do fenômeno.

2) Escolher um número de classes, em geral de mesma amplitude (tamanho), que englobem todos os dados sem haver superposição dos intervalos. Convenciona-se usar de 5 a 20 classes, não há regra rígida para esse número,

mas usa-se também uma fórmula, denominada de "Regra de Sturges" ($NC = 1 + 3{,}3 \, . \log n$).

3) Obter a amplitude de cada classe, dividindo-se a amplitude total pelo valor definido para o número de classes. Esse valor pode ser modificado para facilitar a construção da tabela. Recomenda-se ainda usar um valor redondo.

4) Contar o número de elementos que pertencem a cada classe; este número é denominado freqüência absoluta, em geral representada por f_i

5) Determinar a freqüência relativa de cada classe, dividindo a freqüência da classe pelo número total de observações (N).

EXERCÍCIO RESOLVIDO:

A relação abaixo representa os pesos, em kg, dos 40 alunos de uma turma. Responda às questões de 1 a 9, sabendo que log 40 vale aproximadamente 1,6.

44,5	62	72	72	39,5
41,6	58	73,5	73,5	52,5
40	58,5	42	42	50,4
40	69	58	58	56
42	65	87	47,5	55
55	65	56,5	46,8	65
62	39,5	60	45,7	66,5
63,5	44,5	60	55,9	68

1) De acordo com a fórmula de Sturges, qual o número recomendável de classes para uma distribuição de freqüências contínua?
   NC ≅ 1 + 3,3 x 1,6 ≅ 6,28 ≅ 6 classes.

2) Qual a amplitude total da amostra?
   Será a diferença entre os valores máximo 87 e mínimo 39,5, logo a amplitude é igual a 87 – 39,5 = 47,5

3) Qual a amplitude a ser usada para a distribuição?
   A amplitude que usaremos será 48, para facilitar os cálculos.

4) Qual a amplitude de cada um dos intervalos?
   É só dividir 48 (amplitude total) por 6 (número de classes), logo 48:6 = 8

5) Qual o limite inferior da primeira classe?
   O limite inferior da primeira classe será igual ao menor valor encontrado, ou seja, 39,5.

6) Qual o limite superior da última classe?
   Como fizemos um arredondamento da amplitude total, em meio ponto, teremos que o último limite será igual a 87,5 (maior valor medido, aumentado de 0,5).

7) Construir a tabela da distribuição contínua

Classes (kg)	Freq. (f_i)
39,5 ├── 47,5	12
47,5 ├── 55,5	5
55,5 ├── 63,5	11
63,5 ├── 71,5	7
71,5 ├── 79,5	4
79,5 ├── 87,5	1
$\sum$	40

8) Qual o percentual de pessoas do grupo, cujo peso é inferior a 63,5 kg?

Classes (kg)	Freq. (f_i)	Freq. Relat.	Freq. Relat. acumulada
39,5 ├── 47,5	12	30,0 %	30%
47,5 ├── 55,5	5	12,5 %	42,5%
55,5 ├── 63,5	11	27,5 %	70%
63,5 ├── 71,5	7	17,5 %	87,5%
71,5 ├── 79,5	4	10,0 %	97,5%
79,5 ├── 87,5	1	2,5 %	100%
$\sum$	40	100 %	

Observando a coluna que acrescentamos, com as freqüências relativas, verificamos que o percentual das pessoas com peso inferior a 63,5 kg é igual a 70% que corresponde à freqüência relativa acumulada crescente da terceira classe.

9) Qual o percentual de pessoas do grupo, cujo peso é superior a 71,5 kg?
Agora a resposta será 12,5%, que corresponde à soma das freqüências relativas das duas últimas classes.

10) Qual o percentual de pessoas do grupo, cujo peso é inferior a 58 kg?
Verifique que esse tipo de pergunta (que tem sido bastante cobrada em concursos recentes) é menos óbvia que as anteriores, pelo fato do valor envolvido não ser um dos limites de classe (58 kg). Como devemos proceder num caso desses?

A solução não será difícil, bastando usar o recurso da regra de três, é um método denominado INTERPOLAÇÃO LINEAR PELA OGIVA OU PELAS FREQUÊNCIAS ACUMULADAS CRESCENTES.

Verifique que todas as pessoas inseridas nas duas primeiras classes fazem parte das que atendem ao que se pede na questão (peso inferior a 58 kg). Quanto as que pertencem à terceira classe, nem todas atenderão ao que se pede, já que essa classe vai de 55,5 kg a 63,5 kg. Verificando que a classe toda tem amplitude igual a 8 e freqüência relativa de 17,5%, a amplitude de 2,5 (que vai de 55,5 kg a 58 kg) corresponderá a um percentual que poderemos obter através de regra de três. É claro que é um resultado aproximado, como praticamente tudo o que fazemos nas distribuições de

freqüências, já que sempre que a tabela está pronta, por não dispormos mais das informações detalhadas, há uma perda de informações.

A regra de três que conclui a questão é:

8 ——— 17,5%
2,5 ——— x

$$x = \frac{2{,}5 \text{ x } 17{,}5}{8} \cong 5{,}47\%$$

Conclusão: O percentual das pessoas que têm seu peso inferior a 58 kg é de 30% + 12,5% + 5,47¨% = 47,97% (aproximadamente).

Ao concluirmos nosso capítulo das distribuições de freqüências, cabe destacar ainda que, apesar da simplificação e da agilidade gráfica que possibilitam, tais tabelas contínuas apresentam a desvantagem de **"perda de informações"**.

A título de exemplo, olhando apenas para a tabela acima, do exercício resolvido, tente descobrir quais foram o maior e o menor peso detectados para os 40 alunos da turma?

Vejamos uma questão de concurso recente, sobre o tema:

(Auditor Fiscal da Receita Federal – 2002)
O atributo do tipo contínuo X, observado como um inteiro, numa amostra de tamanho 100 obtida de uma população de 1000 indivíduos, produziu a tabela de freqüências seguinte:

Classes	Freqüência (f)
29,5-39,5	4
39,5-49,5	8
49,5-59,5	14
59,5-69,5	20
69,5-79,5	26
79,5-89,5	18
89,5-99,5	10

Assinale a opção que corresponde à estimativa do número de indivíduos na população com valores do atributo X menores ou iguais a 95,5 e maiores do que 50,5.
a) 700 b) 638 c) 826 d) 995 e) 900

Como a questão envolve agora dois valores e que não são extremos de classes da distribuição, teremos um pouco mais de trabalho, sendo necessário o uso de duas regras de três.
a) Podemos verificar que a terceira classe da tabela terá participação parcial da resposta, pois nem todos os valores de 49,5 a 59,5 são maiores do que 50,5.

Como essa classe (a terceira) tem amplitude 10 e corresponde a uma freqüência absoluta simples de 14 e de 49,5 a 50,5, temos uma amplitude de 1, teremos a seguinte regra de três:

10 -------------- 14
1 ---------------- x

$$x = \frac{1 \text{ x } 14}{10} = 1{,}4$$

b) Verificamos ainda que a quarta classe da tabela entrará totalmente na resposta, já que ela vai de 59,5 a 69,5. Teremos então, para a amostra, mais 20 indivíduos a considerar (freqüência da quarta classe)
c) O mesmo ocorrerá com a quinta classe, que vai de 69,5 a 79,5. Agora serão mais 26 indivíduos a considerar na amostra.
d) Também a sexta classe será considerada na íntegra, pois ela vai de 79,5 a 89.5. Agora temos um acréscimo de mais 18 elementos na resposta
e) Finalmente a última classe da tabela não poderá ser considerada integralmente, pois ela vai de 89,5 a 99,5 e o texto da questão diz que só temos que considerar os valores do atributo X, menores ou iguais a 95,5. Teremos de montar uma outra regra de três, lembrando que a amplitude 10 corresponde a 10 indivíduos e a amplitude 4 (de 89,5 a 95,5) obviamente corresponderá a 4 indivíduos.

Um outro detalhe importante da questão é que a amostra é de 100 pessoas, numa população de 1000 indivíduos. O resultado que estamos obtendo é referente à AMOSTRA, ou seja, a 10% do Universo. Nesse caso, como a pergunta feita foi sobre a população, teremos que multiplicar a nossa resposta por 10 (10% x 10 = 100%).

resposta amostral = 1,4 + 20 + 26 + 18 + 4 = 82,6 elementos.

Resposta populacional = 82,6 x 10 = 826 elementos (opção C)

"Quem quer fazer algo encontra um meio; quem não quer fazer nada encontra uma desculpa".

Provérbio árabe

Exercícios (Lista 3)

21) O Estado de São Paulo apresentou 733 986 matrículas na primeira série , no início do ano de 1973, e 683 816, no fim do ano. O Rio de Janeiro apresentou, respectivamente, 436 127 e 412 457 matrículas. Qual a diferença entre as taxas de evasão escolar dos dois estados nesse ano? (arredondamento a centésimos).
a) 1,41 % b) 1,40 % c) 1,42 % d) 2,45 % e) 3,43 %

22) Um estado brasileiro tinha, em 2002, uma população estimada de 20 636 874 habitantes; sabendo-se que sua área terrestre é de 247 320 km^2, calcule a sua densidade demográfica desse ano.(Arredondamento a centésimos).
a) 84,35 hab/km^2
b) 83,44 hab/ km^2
c) 78,92 hab/ km^2
d) 80,04 hab/ km^2

Uma escola apresentava, no final de um ano, a seguinte tabela de matrículas:

Séries	Março	Novembro
Primeira	480	475
Segunda	458	456
Terceira	436	430
Quarta	420	420
Total	1794	1781

As questões 23, 24 e 25 dizem respeito à tabela acima.

23) Qual a taxa de evasão escolar da primeira série ?
a) 2,34 % b) 4,35 % c) 1,45 % d) 1,04 % e) 3,23 %

24) Qual o coeficiente de evasão escolar, com aproximação a milésimos, da terceira série?
a) 0,014 b) 0,14 c) 0,14 % d) 0,013 e)1,3%

25) Qual a taxa de evasão escolar da escola ?
a) 0,75 % b) 0,71 c) 0,72 % d) 1,2 % e) 0,15 %

26) Qual a amplitude amostral dos dados: 1,2; 3,4 ; 0,9 ; 6,8 ; 1,4 ; 7,8 ; 1,7 .
a) 6,7 b) 9,8 c) 0,5 d) 1,2 e) 6,9

27) Tendo havido uma epidemia na cidade A, com 108 casos em agosto e 12 casos em setembro, certo jornal publicou que houve uma diminuição de 800 % nos casos da doença. Está certa a notícia? Qual foi a redução correta do número de casos da doença?
a) 11 % b) 66 % c) 79 % d) 89 % e) 20%

28) Foi feito um rol com os pesos dos 1000 alunos de um colégio. Aplicando-se a fórmula de Sturges, qual seria o número recomendado de classes para a distribuição de freqüências do tipo contínua?
a) 10 b) 11 c) 12 d) 13 e) 14

29) Em uma distribuição de freqüências, para uma de suas classes obtivemos a freqüência absoluta igual a 12 e a freqüência relativa igual a 0,24. Pode-se dizer que o número de elementos da população observada é:
a) 40 b) 25 c) 100 d) 24 e) 50

30) Qual o número aproximado ideal de classes, para uma distribuição de freqüências onde N = 312.
Aplicar a relação de Sturges, sabendo-se que log 312 vale aproximadamente 2,5.
a) 9 b) 8 c) 11 d) 10 e) 6

31) Observe a tabela:

Unidades Escolares (Ensino Fundamental)

Anos	**Quantidade**
2002	**189 900**
2003	**190 345**
2004	**195 400**
2005	**198 600**

Fonte: Dados Fictícios

Esta tabela é uma série do tipo:
a) específica b) conjugada c) geográfica d) cronológica e) mista

32) Um conjunto de 500 notas de Estatística, extraídas dos arquivos da secretaria de um Colégio, constitui:
a) Um rol
b) uma relação de dados brutos
c) uma tabela
d) uma distribuição de freqüências
e) um gráfico estatístico

33) As regras básicas para se construir uma distribuição de freqüências contínuas são:
I. Nenhum dado deve ser excluído
II. Nenhum dado deve ser computado mais de uma vez.
III. As classes devem ser mutuamente exclusivas.
IV. campo de variação da variável deve ser completamente esgotado.

Destas regras:
a) todas estão corretas
b) todas estão erradas.
c) só a segunda está errada.
d) só a terceira está errada.
e) só a quarta está correta.

34) Quando afirmamos que as classes de uma distribuição de freqüências devem ser mutuamente exclusivas, estamos querendo que:

a) Nenhum dado seja excluído
b) Nenhum dado seja contado mais de uma vez.
c) todos os dados sejam computados.
d) possamos exaurir totalmente o campo de variação da variável.
e) Os limites de classe sejam considerados.

35) Os elementos fundamentais que o título de uma tabela deve conter são:
a) veracidade, local e fonte
b) fato, local e época
c) veracidade , clareza e fato
d) época, fonte e clareza.
e) fonte, fato, local e época.

As questões de 36 a 42 referem-se aos conceitos dos elementos de uma distribuição de freqüências e as simbologias usadas. Selecione, para cada questão, a alternativa correta.

a)	**freqüência relativa**
b)	**freqüência absoluta**
c)	**limite de intervalo**
d)	**ponto médio**
e)	**população observada**

36) É o ponto central de um intervalo de classe:

37) É o valor do extremo superior de um intervalo:

38) É o limite inferior do intervalo, acrescido da metade da amplitude do mesmo:

39) Simbologia: $\mathbf{f / \sum f}$

40) $\mathbf{(\ell_i + L_i) : 2}$

41) Número de observações de cada dado:

42) Simbologia: $\mathbf{\sum f_i}$

As questões de 43 a 45, dizem respeito à tabela seguinte:

Pontos	Alunos
0|-------- 5	**02**
5|-------- 10	**05**
10|-------- 15	**15**
15|-------- 20	**08**
Total	**30**

43) Qual a freqüência relativa, correspondente ao segundo intervalo?
a) 0,15 b) 0,16 c) 0,17 d) 0,18 e) 0,19

44) Qual o ponto médio do último intervalo de classe?
a) 15,0 b) 17,5 c) 20,0 d) 22,5 e) 12,5

45) Qual a amplitude da distribuição:
a) 10 b) 5 c) 15 d) 20 e) 24

46) A primeira etapa, de modo a se construir uma distribuição de freqüência contínua, consiste em calcular:
a) a amplitude do intervalo de classe
b) o limite inferior da primeira classe
c) a amplitude total da distribuição
d) o limite superior da última classe

47) A freqüência relativa é obtida:
a) adicionando-se a freqüência absoluta ao somatório das freqüências posteriores
b) dividindo-se a freqüência absoluta pelo produtório das freqüências.
c) dividindo-se por dois o somatório das freqüências e multiplicando por 100.
d) dividindo-se a freqüência absoluta pela freqüência acumulada crescente da última classe.
e) nada disso

As questões de 48 a 50 dizem respeito à tabela de preços abaixo, que representa a pesquisa em 20 lojas , referente a determinado produto:

Preços em Reais	**Número de lojas**
50	2
51	1
52	5
54	6
55	2
56	3
60	1

48) Qual o percentual de lojas que estão cobrando até R$52,00 (inclusive)?
a) 45 % b) 40 % c) 36 % d) 50 % e) 60 %

49) Qual o percentual de lojas que estão cobrando acima de R$ 55,00?
a) 15 % b) 30 % c) 40 % d) 25 % e) 20 %

50) Qual a freqüência relativa do preço mais caro da tabela?
a) 2% b) 3 % c) 4 % d) 5 % e) 6 %

"Não se pode ensinar tudo a alguém,
pode-se apenas ajudá-lo a encontrar por si mesmo."

Galileu Galilei, astrônomo italiano

51) (Concurso TTN - 1994)

Considere a distribuição de freqüência transcrita a seguir:

Peso (Kg)	Freq. Simples
2 |-------- 4	9
4 |-------- 6	12
6 |-------- 8	6
8 |--------10	2
10|--------12	1

a) 65% das observações têm peso não inferior a 4 kg e inferior a 10 kg.
b) Mais de 65 % das observações têm peso maior ou igual a 4 kg.
c) Menos de 20 observações têm peso igual ou superior a 4 kg.
d) A soma dos pontos médios dos intervalos de classe é inferior ao tamanho da população.
e) 8% das observações têm peso no intervalo de classe 8 |____10

As questões 52, 53 e 54 dizem respeito à tabela abaixo:

Número de ítens	Número de pedidos
10 |--------- 15	3
15 |--------- 20	7
20 |---------25	16
25 |---------30	12
30 |---------35	9
35 |---------40	5
40 |---------45	2
Total	54

52) Assinale a alternativa errada:
a) o número de elementos é 7.
b) o limite inferior de segunda categoria é 15.
c) o intervalo de classe é 5
d) o ponto médio da terceira categoria é 22,5.
e) o número de elementos é 54.

53) Essa tabela é classificada como sendo de:
a) dupla entrada ou bidimensional.
b) simples ou unidimensional.
c) cruzada.
d) cronológica.
e) geográfica.

54) Qual a porcentagem de pedidos abaixo de 25 itens?
a) 48% b) 54% c) 45% d) 65% e) 25%

55) Complete a distribuição de freqüências abaixo e determine a freqüência relativa da última classe (%).

CLASSES F_i	F_{ac}	
0 ________| 2	3	
2 ________| 4	**?**	8
4 ________| 6	8	
6 ________| 8	10	
8 ________|10		28

a) 7,14% b) 7,18% c) 6,47% d) 5,68% e) 7,23%

GABARITO

21) A	**22) B**	**23) D**	**24) A**	**25) C**
26) E	**27) D**	**28) B**	**29) E**	**30) A**
31) D	**32) B**	**33) A**	**34) B**	**35) B**
36) D	**37) C**	**38) D**	**39) A**	**40) D**
41) B	**42) E**	**43) C**	**44) B**	**45) D**
46) C	**47) D**	**48) B**	**49) E**	**50) D**
51) B	**52) A**	**53) B**	**54) A**	**55) A**

"Somos o que fazemos, mas somos, principalmente,
o que fazemos para mudar o que somos."
(Eduardo Galeano)