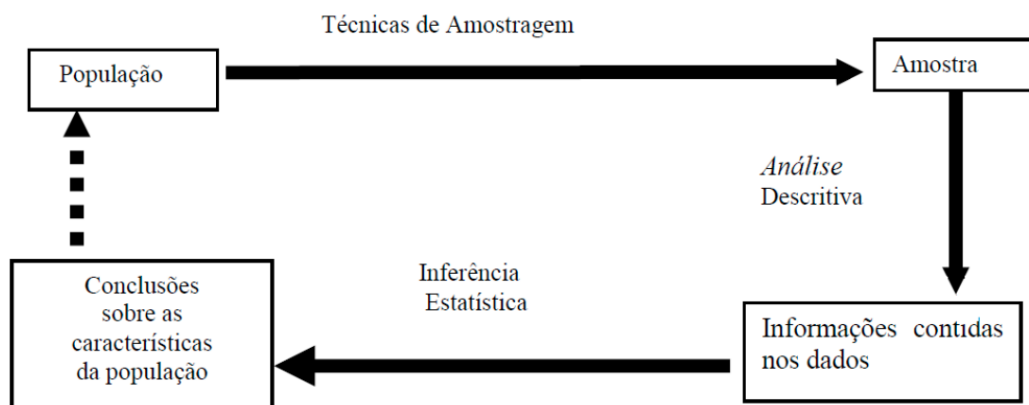


Objetivos

- conceitos básicos de população e amostra de uma pesquisa estatística;
- distinção entre variáveis qualitativas e variáveis quantitativas;
- construir distribuições de frequências para var. qualitativas e quantitativas discretas;
- construir gráficos de colunas e de setores p/ representar dados qualitativos e quantitativos discretos.

Conceitos

- **Estatística** é a ciência da aprendizagem a partir dos dados.
- Ela constitui-se num **conjunto de técnicas e métodos científicos** que tratam da coleta, análise e interpretação de informações numéricas, cujo objetivo principal é auxiliar na tomada de decisões ou tirar conclusões em situações de incerteza, a partir de informações numéricas.
- A Teoria Estatística moderna se divide em dois grandes campos:
 - **Estatística Descritiva**
um conjunto de métodos que ensinam a reduzir uma quantidade de dados bastante numerosa por um número pequeno de medidas, substitutas e representantes daquela massa de dados.
 - **Estatística Indutiva ou Inferência Estatística**
consiste em inferir (deduzir ou tirar conclusões a respeito das) propriedades de um universo a partir de uma amostra. O processo de generalização, que é característico do método indutivo, está associado a uma margem de incerteza. A medida da incerteza é tratada mediante técnicas e métodos que se fundamentam na Teoria das Probabilidades.



- **População (ou universo de interesse)**: Conjunto de indivíduos, objetos ou informações que apresentam pelo menos uma característica comum, cujo comportamento interessa-nos analisar.
Ou, em outras palavras:
 - conjunto de todas as medidas, observações relativas ao estudo de determinado fenômeno, ou;
 - conjunto de elementos para os quais se deseja estudar determinada(s) característica(s).
- **Ex 1**: um banco querendo lançar um novo produto, precisa conhecer o perfil socioeconômico dos seus clientes
População: todos os seus clientes
Característica: X = perfil sócioeconômico
- **Ex 2**: estudo sobre a precipitação pluviométrica na Região Nordeste no ano 2000
População ou universo: área referente à Região Nordeste
Característica: X = precipitação pluviométrica
- **OBS**: uma população pode ser **finita** (quantidade de cursos na UNEB) ou **infinita** (totalidade de grãos de areia no globo terrestre)
- **OBS**: o estudo de toda uma população (**censo**) embora forneça dados completos e precisos, muitas vezes é inviável devido a custos elevados, tempo demandado e complexidade logística ou, impossível se a população for infinita. Assim, estudar **parte** da população constitui-se um aspecto fundamental da Estatística.

- **Amostra: (parte representativa do todo** – Diante das dificuldades de se estudar uma população inteira)
 - A grosso modo, é qualquer subconjunto da população;
 - É um subconjunto, uma parcela selecionada da população, que deve ser representativa do todo;
 - O objetivo é que as características observadas na amostra possam ser generalizadas para toda a população da qual ela foi extraída.
 - **Resumo: população** é o grupo completo sobre o qual se deseja tirar conclusões, enquanto a **amostra** é o grupo menor que é efetivamente estudado para inferir informações sobre a população.
 - **OBS:** A validade de qualquer pesquisa estatística baseada em amostragem depende diretamente da qualidade e representatividade da amostra escolhida.

Cenário da Pesquisa	População	Amostra
Intenção de voto para presidente do Brasil	Todos os eleitores aptos a votar no Brasil	Um grupo de 5.000 eleitores de diferentes estados, faixas etárias e classes sociais
Qualidade dos parafusos produzidos em uma fábrica	Todos os parafusos produzidos pela fábrica em um determinado período	Uma caixa com 200 parafusos selecionados aleatoriamente da linha de produção
Estudo sobre os hábitos de leitura dos estudantes de uma cidade	Todos os estudantes matriculados nas escolas daquela cidade	500 estudantes selecionados de diferentes escolas, séries e bairros da cidade.

- **Tipos de Variáveis**
 - As características da população que nos interessa analisar recebem o nome de **variáveis**.
 - Podem ser divididas em dois tipos: **qualitativas** e **quantitativas**.
- **Variáveis qualitativas:** quando o resultado da observação é apresentado na forma de qualidade ou atributo.
 - Ex 3:** sexo; estado civil; grau de escolaridade; classe social, nacionalidade, gênero, etc
 - OBS:** São divididas em dois tipos:
 - Nominal** (não existe nenhuma ordenação nas possíveis respostas da referida variável)
 - EX 4:** sexo, cor dos olhos, estado civil, tipo sanguíneo, nacionalidade, gênero, nacionalidade, etc
 - ordinal** (para a qual existe uma ordenação)
 - EX 5:** classe social, grau de instrução, porte de empresa, grau de uma doença, etc
- **Variáveis quantitativas:** quando o resultado da observação é um número, decorrente de um processo de mensuração ou contagem.
 - **Ex 6:** número de filhos; salário mensal; altura; peso; idade; tamanho da família; etc.
 - **OBS:** são divididas em:
 - discretas** (que assumem valores em um conjunto finito ou enumerável de números)
 - Ex 7:** número de filhos, número de carros, número de defeitos...
 - contínuas** (que assumem valores em um intervalo números reais)
 - Ex 8:** peso, altura, vida útil de bateria...

Atividade 1

Identifique as variáveis pesquisadas, classificando-as como qualitativas ou quantitativas.

“O Levantamento Socioeconômico (LSE) é a pesquisa do IBOPE Mídia que mapeia as características sociais, demográficas e econômicas das famílias das principais regiões metropolitanas do país. Oferece também outros dados essenciais para traçar a estratégia de marketing para um produto. Com uma base de dados estendida em relação às outras pesquisas do IBOPE Mídia, o LSE serve de base para outros estudos.

São levantados dados sobre a condição do domicílio entrevistado (condição da rua, tipo de imóvel) e sobre a condição socioeconômica do domicílio (informações sobre renda e classificação econômica). Também são pesquisados o número de pessoas no domicílio, a presença e a quantidade de crianças e adolescentes, a idade, grau de instrução e condição de atividade do chefe da casa e da dona-de-casa. A pesquisa levanta também dados sobre a posse de bens, como geladeira, máquina de lavar, automóvel, rádio, computador, telefone, entre outros, e acesso a serviços de mídia, como TV por Assinatura, Internet, etc.”

Fonte: www.ibope.com.br (parte da descrição da pesquisa sociodemográfica)

Resposta atividade 1:

Condição do domicílio	variável qualitativa
Condição da rua	variável qualitativa
Tipo de imóvel	variável qualitativa
Renda	Qualitativa, se for perguntada a faixa ou quantitativa, se for perguntada a renda exata; a 1a opção é a mais provável para esse tipo de pesquisa
Classificação econômica	variável qualitativa
Número de pessoas	variável quantitativa
Presença de crianças	variável qualitativa
Número de crianças	variável quantitativa discreta
Presença de adolescentes	variável qualitativa
Número de adolescentes	variável quantitativa discreta
Idade do chefe e da dona-de-casa	pode ser quantitativa, caso se pergunte a idade exata, ou qualitativa, caso se identifique a faixa etária
Grau de instrução do chefe e da dona-de-casa	variável qualitativa
Condição de atividade do chefe	variável qualitativa
Presença de geladeira, máquina de lavar, etc.	variáveis qualitativas do tipo Sim/Não
Acesso a serviços de mídia	variáveis qualitativas do tipo Sim/Não

Atividade 2

Os dados abaixo (fictícios) representam o rendimento em porcentagem de uma reação para fabricação de uma substância química, em 80 bateladas produzidas por uma indústria. A empresa decidiu construir tabela(s) e gráfico(s) para obter um resumo do conjunto de dados.

70,7 71,8 73,9 74,4 75,9 76,0 76,6 76,7 77,4 78,0 78,1 78,1 78,2 78,4 78,4 78,4
78,5 78,5 78,5 78,9 79,0 79,1 79,3 79,3 79,5 79,5 79,7 79,8 79,9 79,9 80,1 80,2
80,4 80,4 80,5 80,7 80,7 80,7 80,9 81,3 81,4 81,6 81,8 81,9 82,0 82,0 82,1 82,3
82,5 82,7 82,9 83,0 83,0 83,2 83,4 83,5 83,6 83,6 83,7 83,8 84,3 84,5 84,5 84,5
84,6 85,2 85,5 85,5 85,7 86,4 86,5 86,8 86,8 86,8 87,1 87,1 87,1 87,3 88,5 90,0

○ **Apresentação dos dados:**

- procedimentos para organizar e descrever um conjunto de dados (numa população ou amostra): **tabelas** ou **gráficos**.
- o conjunto de informações disponíveis, após a tabulação do questionário ou pesquisa de campo, é denominado de **planilha de dados** (ou **tabela de dados brutos**), sem qq organização numérica.
- Quando é organizada em alguma ordem, chamamos de **rol**.

Ex 9: Na Região XYZ foi realizado um levantamento para investigar algumas característica das empresas no ano 2 e o faturamento nos anos 1 e 2. O banco de dados foi organizado conforme o seguinte esquema:

- Identificação da empresa (**empresa**) – número sequencial
- Setor de atividade industrial (**setorind**) – por códigos: Mobiliário=0; bebidas=1; Editorial=2; Vestuário=3; Minerais=4; Produtos alimentares=5
- Faturamento do ano 1 (**fat_ano1**) – valores em mil unidades monetárias
- Faturamento do ano 2 (**fat_ano2**) – valores em mil unidades monetárias
- Número de empregados (**emprego**)
- Tempo de funcionamento da empresa (**idade**) – valores em anos completos

empresa	setorind	fat_ano1	fat_ano2	emprego	idade

Ex 10: A direção de uma empresa está estudando a possibilidade de fazer um seguro saúde para seus funcionários e respectivos familiares. Para isso, ela faz um levantamento junto a seus 500 funcionários, obtendo informação sobre **sexo, estado civil, idade, número de dependentes e salário**.

*Como pode ser organizada a **planilha de dados brutos**?*

	func	est_civil	idade	n_dep	salario
1					
2					
...					
500					

OBS: Todas as variáveis podem ser resumidas através de uma tabela, mas a construção é diferenciada dependendo do tipo de variável.

- **Tabela Simples:** apresenta as respectivas contagens (frequência de ocorrência);
- **Distribuição de Frequência:** resumo dos valores coletados em classes, categorias ou intervalos convenientemente estabelecidos.

OBS: tabela de freq. p/ variável quant. discreta = distribuição de freq. p/ dados **não-agrupados em classes**

○ **Tabelas**

A apresentação através de tabelas, em uma pesquisa, segue algumas convenções/normas, dependendo de qual instituição, congresso ou órgão, esta tabela será apresentada. Mas alguns princípios básicos, podemos utilizar, segundo as normas do IBGE (Instituto Brasileiro de Geografia e Estatística):

- **Título:** aonde é dada uma noção inicial ao leitor sobre o que é a tabela;
- **Cabeçalho:** para que sejam identificados os conteúdos referentes a cada coluna da tabela. O cabeçalho deve conter o suficiente para responder as questões:
o que está sendo representado? onde ocorreu? Quando ocorreu ?
- **Coluna Indicadora:** que especifica as diferentes categorias da variável;
- **Corpo:** é representado por colunas e subcolunas dos quais são registradas as informações.
- **Rodapé ou pé:** onde é identificada a fonte original dos dados, ou alguma nota referente a tabela.

Título

Cabeçalho

Corpo

Rodapé

Tabela 1: Distribuição do número de funcionários por sexo	
Sexo	Frequência Simples
Masculino	270
Feminino	230
Total	500

Fonte: censo interno

- **Frequência (simples) absoluta** é o número de vezes que um determinado valor da variável aparece
- **Frequência relativa** é a frequência de cada categoria em relação ao total:

Tabela 2: Distribuição do número de funcionários por sexo

Sexo	Frequência	
	Absoluta	Relativa (%)
Masculino	270	54
Feminino	230	46
Total	500	100

Fonte: censo interno

Tabela 3: Distribuição do número de funcionários por estado civil

Estado Civil	Frequência	
	Absoluta	Relativa (%)
Solteiro	125	25
Casado	280	56
Divorciado	85	17
Viúvo	10	2
Total	500	100

Fonte: censo interno

OBS: A soma das frequências relativas é sempre 1 (ou 100%), enquanto a soma das frequências absolutas deve ser igual ao número total de elementos sendo investigados.

EX 11: Consideremos que, na situação do **EX 10**, os dados tenham sido levantados por departamento, para depois serem totalizados. Para o Departamento de RH, foram obtidas as seguintes informações:

Tabela 4: Distribuição dos funcionários do RH

	Nome	Sexo	Estado Civil	Número de dependentes
1	João da Sila	M	casado	3
2	Pedro Gomes	M	viúvo	1
3	Maria Freitas	F	casada	0
4	Paula Dantas	F	solteira	0
5	Ana Soares	F	solteira	1
6	Luiz Costa	M	casado	3
7	André Souza	M	casado	4
8	Patrícia Silva	F	divorciada	2
9	Regina Lima	F	casada	2
10	Alfredo Souza	M	casado	3
11	Rita Cunha	F	solteira	0
12	Pedro Bastos	M	divorciado	2
13	Ricardo Silva	M	solteiro	0
14	Márcio Rezende	M	solteiro	1
15	Joana Soares	M	solteira	0

Fonte: censo interno

Tabela 5: Distribuição do número de funcionários do RH por sexo

Sexo	Frequência	
	Absoluta	Relativa %
Masculino	8	53,33
Feminino	7	46,67
Total	15	100

Fonte: censo interno

Tabela 6: Distribuição do número de funcionários do RH por estado civil

Estado Civil	Frequência	
	Absoluta	Relativa %
Solteiro	6	40,00
Casado	6	40,00
Divorciado	2	13,33
Viúvo	1	6,67
Total	15	100

Fonte: censo interno

Frequência acumulada:

é o somatório das frequências de todas as classes de uma tabela, incluindo a frequência da classe em questão e todas as classes anteriores a ela

Atividade 3: com as informações sobre o tempo de existência das empresas (idade) na Região XYZ organize os dados, construindo: (a) o rol; (b) tabela de frequência.

Tabela 7: Distribuição do número de dependentes

Número de dependentes	Frequência Simples		Frequência Acumulada	
	Absoluta	Relativa	Absoluta	Relativa
0	120	24,0	120	24,0
1	95	19,0	215	43,0
2	90	18,0	305	61,0
3	95	19,0	400	80,0
4	35	7,0	435	87,0
5	30	6,0	465	93,0
6	20	4,0	485	97,0
7	15	3,0	500	100,0
Total	500	100		

Dados Brutos das empresas do EX 9 (Região XYZ)

empresa	setorind	fat_ano1	fat_ano2	emprego	idade	empresa	setorind	fat_ano1	fat_ano2	emprego	idade
1	2	70	45	13	10	63	4	446	381	18	13
2	2	48	65	13	16	64	4	444	381	14	20
3	2	80	66	16	8	65	3	328	386	19	4
4	2	62	69	10	10	66	3	343	389	20	1
5	2	69	77	6	4	67	3	352	391	16	5
6	2	80	88	12	8	68	3	369	402	20	4
7	2	74	89	12	14	69	4	468	402	8	20
8	2	93	98	10	8	70	3	352	405	26	7
9	2	96	104	10	2	71	4	450	410	15	18
10	2	133	122	18	11	72	4	494	415	12	21
11	2	254	123	5	18	73	4	488	423	11	16
12	2	95	129	5	3	74	3	383	426	21	5
13	2	163	145	9	4	75	3	360	432	6	4
14	4	140	150	13	15	76	3	373	432	19	5
15	2	116	154	12	7	77	4	494	432	10	17
16	2	160	154	13	9	78	3	353	450	19	1
17	0	76	156	3	11	79	3	378	455	9	3
18	1	148	156	5	20	80	4	537	455	22	18
19	4	148	167	15	14	81	3	397	456	17	6
20	2	180	171	14	13	82	3	407	463	14	5
21	1	202	174	12	13	83	3	443	465	16	7
22	1	124	176	11	5	84	3	411	473	12	6
23	1	154	183	10	12	85	4	550	475	6	19
24	3	155	185	21	6	86	3	428	483	22	6
25	0	120	192	5	5	87	3	422	483	20	7
26	3	174	202	17	4	88	3	437	487	23	5
27	3	437	205	23	3	89	5	446	487	22	15
28	2	202	212	12	14	90	3	425	489	21	8
29	3	190	217	18	5	91	4	576	490	17	17
30	4	202	230	15	21	92	4	573	495	11	13
31	3	205	233	26	3	93	4	594	500	9	14
32	1	249	233	15	8	94	3	455	502	24	5
33	3	218	236	16	4	95	3	469	503	21	6
34	3	273	245	13	5	96	3	446	505	18	5
35	3	216	248	18	5	97	4	579	510	8	15
36	3	223	250	18	5	98	4	499	510	13	18
37	3	205	253	13	5	99	3	445	525	15	3
38	5	203	254	15	6	100	3	374	530	27	5
39	3	225	261	19	5	101	5	510	534	31	10
40	4	308	265	7	16	102	4	774	543	17	30
41	3	226	268	13	6	103	4	633	545	15	13
42	3	260	276	17	5	104	4	653	567	16	18
43	4	329	282	12	16	105	3	513	570	20	5
44	1	336	292	10	15	106	4	590	585	20	20
45	3	236	299	19	6	107	4	716	610	11	23
46	3	262	310	15	6	108	5	554	620	23	12
47	5	271	310	17	7	109	4	719	623	20	18
48	3	270	314	21	5	110	4	764	630	20	21
49	3	304	333	12	8	111	4	650	634	11	24
50	4	322	334	27	16	112	5	656	670	16	12
51	3	297	335	18	4	113	4	807	682	20	24
52	5	222	345	15	7	114	5	686	695	19	6
53	3	318	351	20	6	115	4	813	702	13	10
54	3	313	354	23	2	116	4	711	723	14	17
55	4	425	355	16	18	117	4	841	730	14	17
56	4	422	361	15	19	118	4	824	734	18	21
57	3	296	365	15	6	119	4	850	750	17	25
58	3	307	367	25	6	120	4	959	810	18	18
59	3	319	369	16	5	121	5	810	835	26	5
60	4	435	370	15	17	122	4	866	875	12	23
61	4	436	372	12	19	123	5	1150	1243	39	11
62	4	453	375	5	12	124	5	1554	1345	29	10

- **Distribuição (ou tabela) de frequência com variável contínua**
As cada