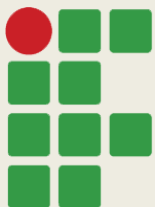


Probabilidade e Estatística



INSTITUTO FEDERAL

Catarinense

Campus Blumenau

Professor Jeovani Schmitt



Probabilidade e Estatística

Aula 2

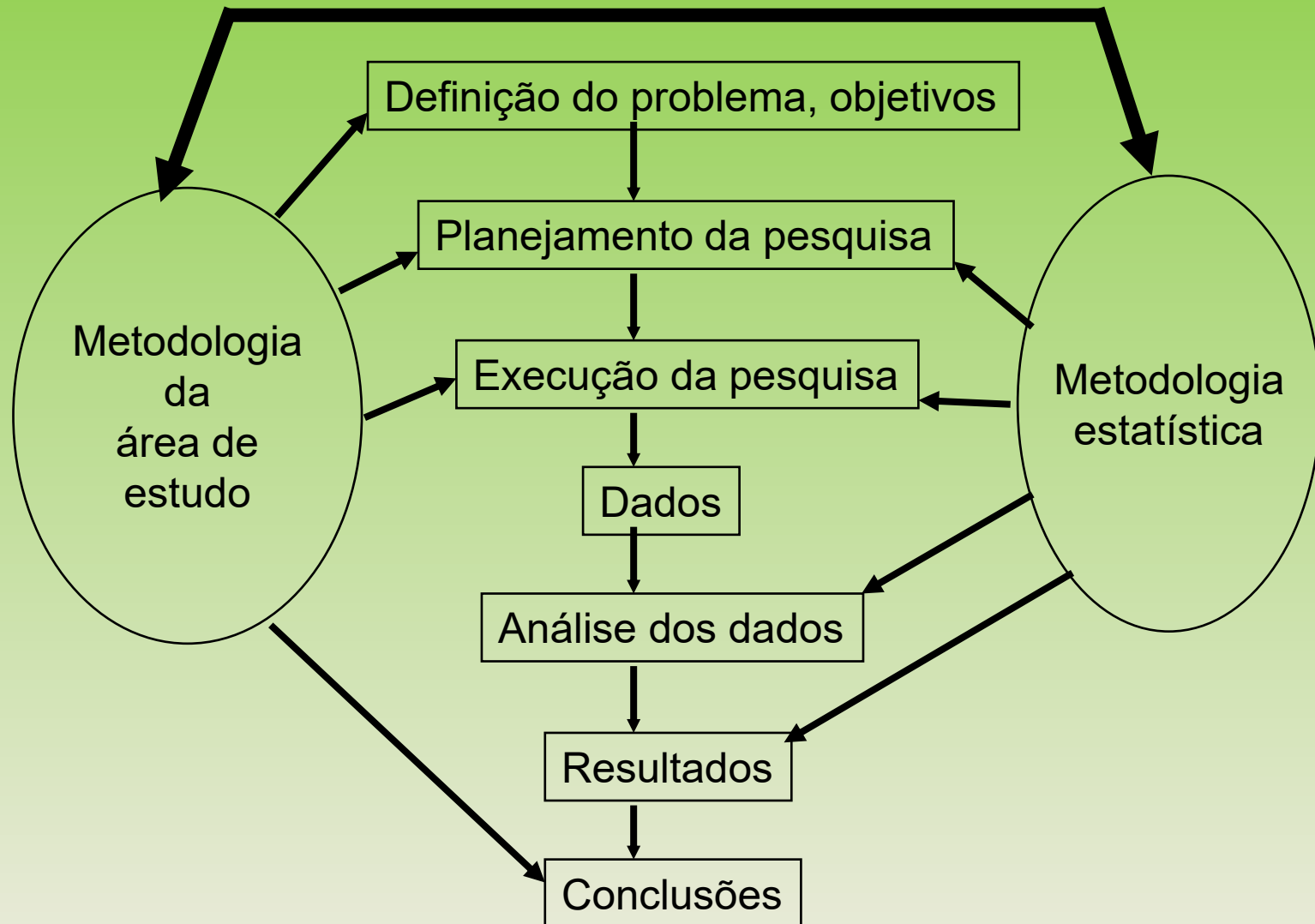
Conceitos básicos

- ✓ Planejamento de uma pesquisa
- ✓ Pesquisas de levantamento
- ✓ Técnicas de amostragem
- ✓ Tamanho da amostra



Capítulo 2

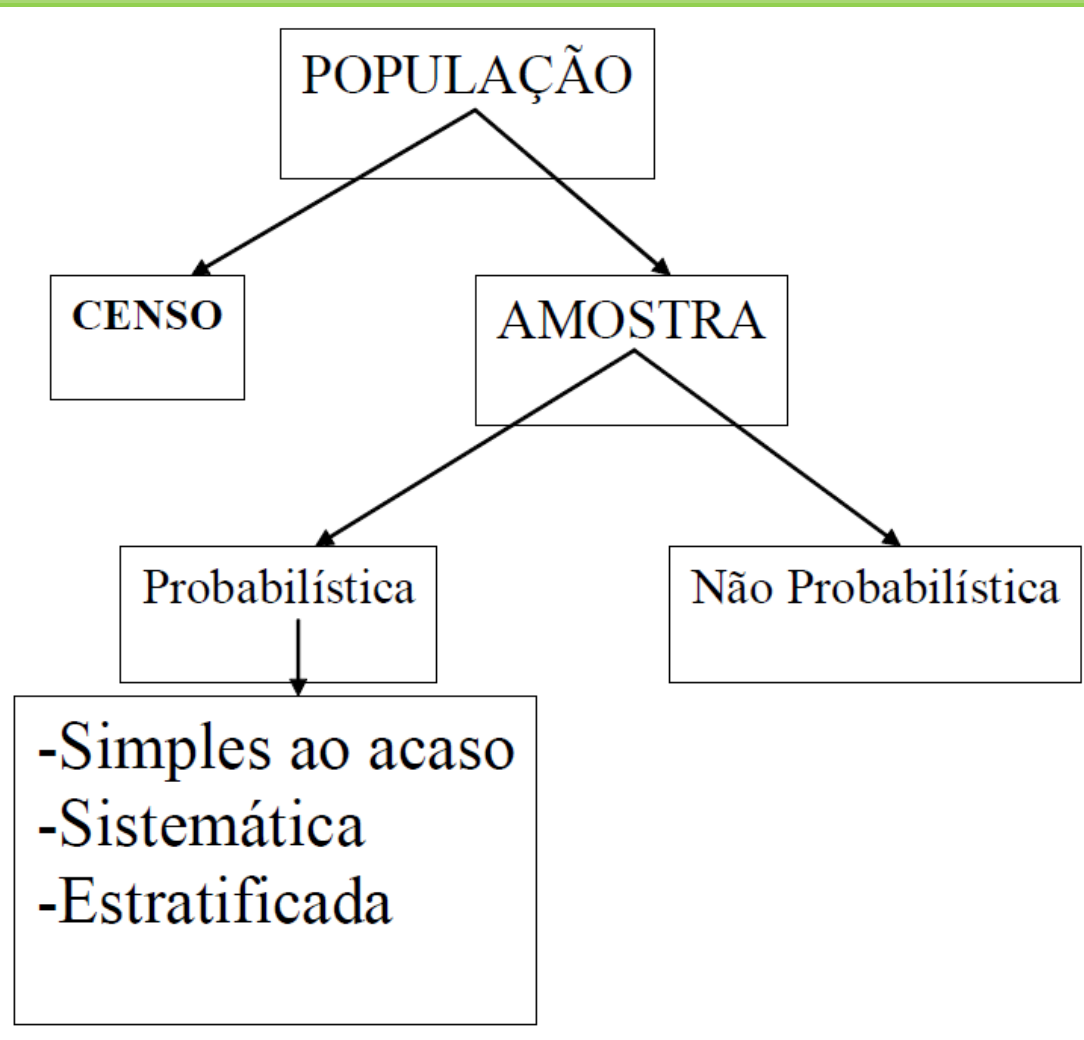
Planejamento de uma Pesquisa



Algumas questões importantes:

- ✓ Quem é a População alvo?
- ✓ Dados primários ou secundários?
- ✓ Quais as variáveis a serem levantadas?
- ✓ Qual será o instrumentos para a mensuração das variáveis?
- ✓ Execução da Pesquisa: Censo ou Amostragem?

Execução da pesquisa



Coleta: dados primários ou secundários?

• Dados secundários

Dados já existentes que podem ser obtidos através de uma publicação, bases de dados disponíveis na internet, sendo assim não precisamos ir até os elementos da população.

Diversas são as fontes de dados que podem ser utilizadas dependendo da pesquisa realizada.

Exemplos:

Educação – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep)

Espaço – Instituto Nacional de Pesquisas Espaciais

Meio ambiente - Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis (IBAMA)

Economia – Instituto de Pesquisa Econômica Aplicada (IPEA)

Estatísticas – Instituto Brasileiro de Geografia e Estatística (IBGE)

Desenvolvimento Humano – Atlas Brasil

Coleta: dados primários ou secundários?

- ## Dados secundários

Exemplos: Dados sobre Educação no INEP

<https://www.gov.br/inep/pt-br>

The screenshot shows the gov.br website header with the logo and navigation links: Ministério da Educação, Órgãos do Governo, Acesso à Informação, Legislação, and Acessibilidade. Below the header, the name of the institution, Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira | Inep, is displayed. A search bar on the right contains the text 'O que voc...'. The main navigation menu is open, showing a list of categories on the left and a list of sub-items on the right. The 'Acesso à Informação' category is selected, and the 'Dados Abertos' sub-item is highlighted.

Acesso à Informação	
Áreas de Atuação	Ações e Programas
Assuntos	Agenda de Autoridades
Canais de Atendimento	Auditorias
Centrais de Conteúdo	Convênios e Transferências
Composição	Dados Abertos
	Informações Classificadas
	Institucional

Indicadores Educacionais	
	Inep Data
	Microdados
	Plano de Dados Abertos
	Serviço de Acesso a Dados Protegidos
	Sinopses Estatísticas

Dados secundários

Datasets

Explore, analyze, and share quality data. [Learn more](#) about data types, creating, and collaborating.

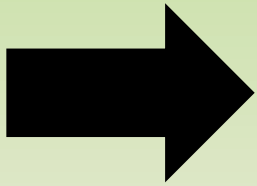
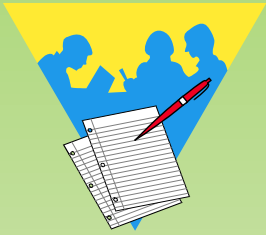


<https://www.kaggle.com/datasets>

Coleta de dados

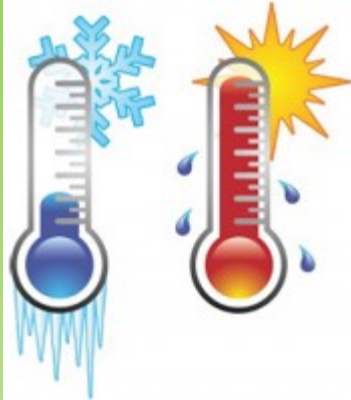
- **Dados Primários**

Necessitam serem coletados observando diretamente cada elemento.

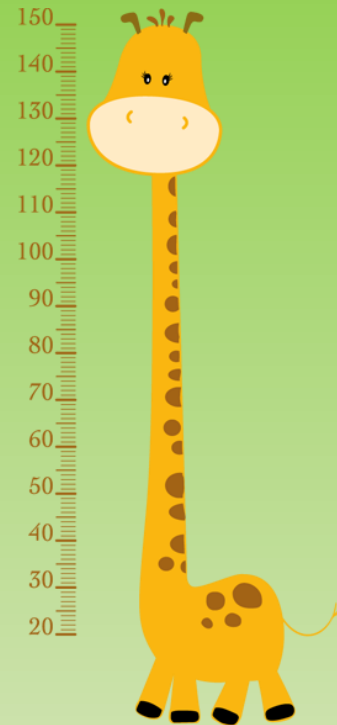


Requer instrumento de mensuração

Instrumento para a mensuração de variáveis



Multímetro
Alicate amperímetro



- Questionário
- Variáveis em seres humanos

o instrumento de pesquisa
está intrinsecamente
relacionado às
variáveis da pesquisa

Coleta de dados - Aplicação

- **Dados Primários**

Exemplo: O questionário da turma representa uma coleta de dados primários.

Questionário

1 – Seu curso?

() BCC

() BEE

2 – Peso: _____ kg

3 – Altura: _____ metros

4 – Idade: _____ anos

5 – Quantos irmãos você tem? _____

6 – Cidade onde você mora?

7 – Você trabalha?

() Sim

() Não

Como se obtém uma amostra?

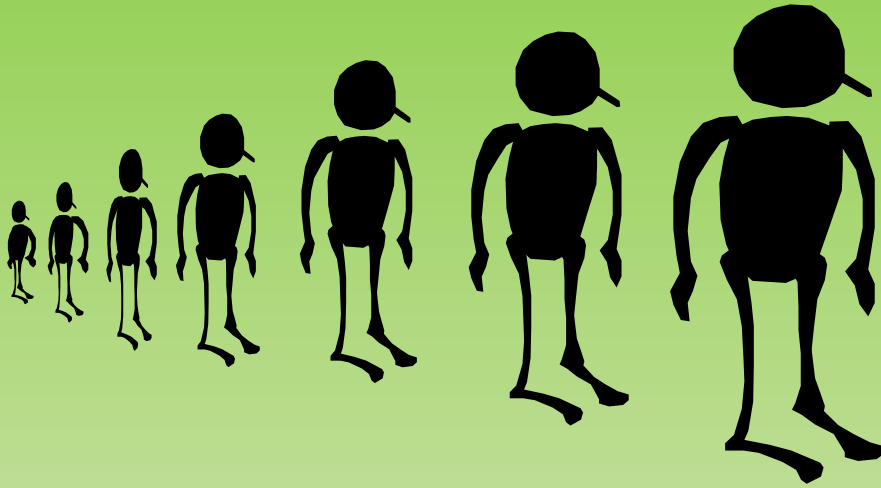
Quantos elementos? (Qual o tamanho da amostra?)

Técnicas de amostragem

1. Amostragem simples ao acaso (ASA)
2. Amostragem sistemática (AS)
3. Amostragem estratificada (AE)
4. Amostragem em estágios (AC – conglomerados)
5. Amostragem não probabilística

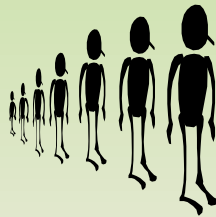
IMPORTANTE! Somente em uma amostragem probabilística é que podemos generalizar os resultados para uma população! Isto é, associar aos resultados uma probabilidade de que estejam corretos, ou seja uma medida da confiabilidade das conclusões obtidas para uma inferência sobre a população.

Amostragem Simples ao Acaso (ASA)



População homogênea
em relação à variável
de interesse!

Sorteio não viciado

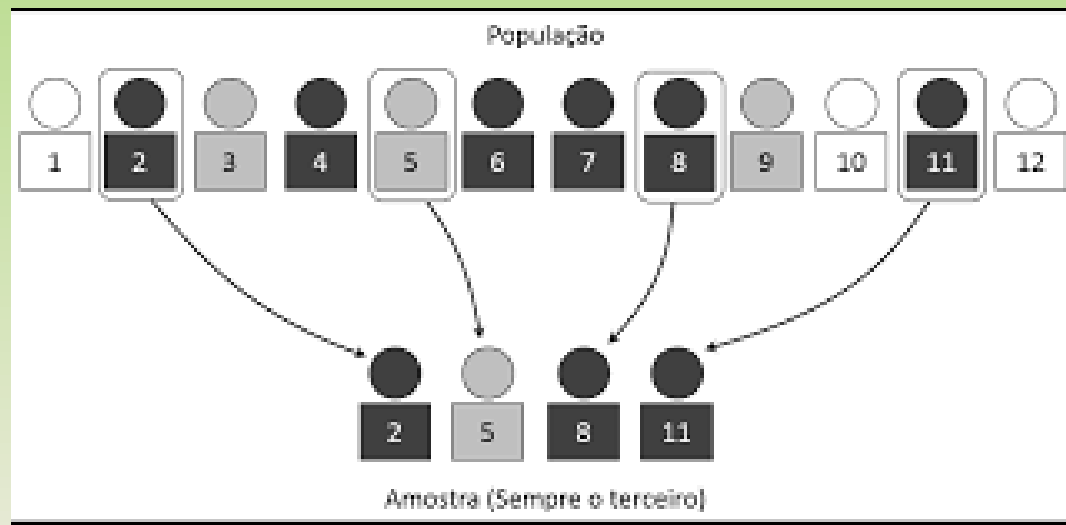


Amostra

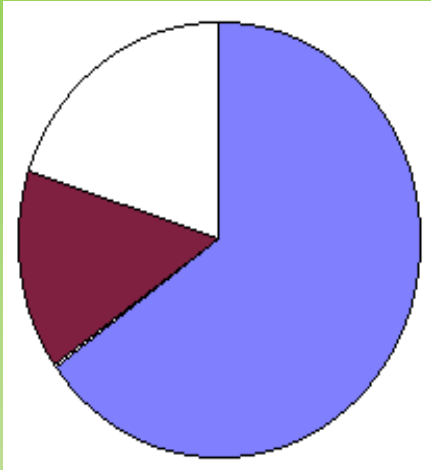
Números aleatórios ou
pseudo-aleatórios

Amostragem Sistemática (AS)

- A listagem é ORDENADA.
 - Divide-se o tamanho da população (N) pelo tamanho da amostra (n), obtendo um intervalo de retirada ($k=N/n$).
 - Sorteia-se o ponto de partida.
 - A cada k elementos retira-se um para a amostra.



Amostragem Estratificada (AE)



População HETEROGÊNEA.

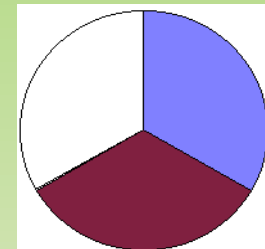
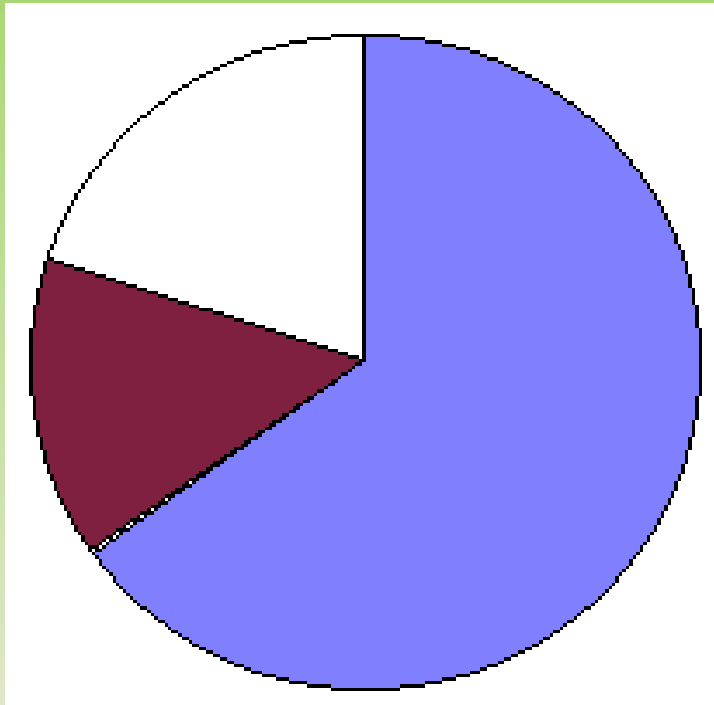
Homogeneidade DENTRO de cada estrato.

Sorteio não viciado

TODOS os estratos precisam ser representados na amostra!

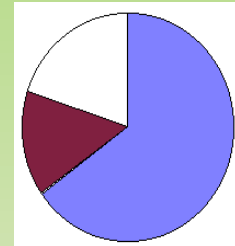
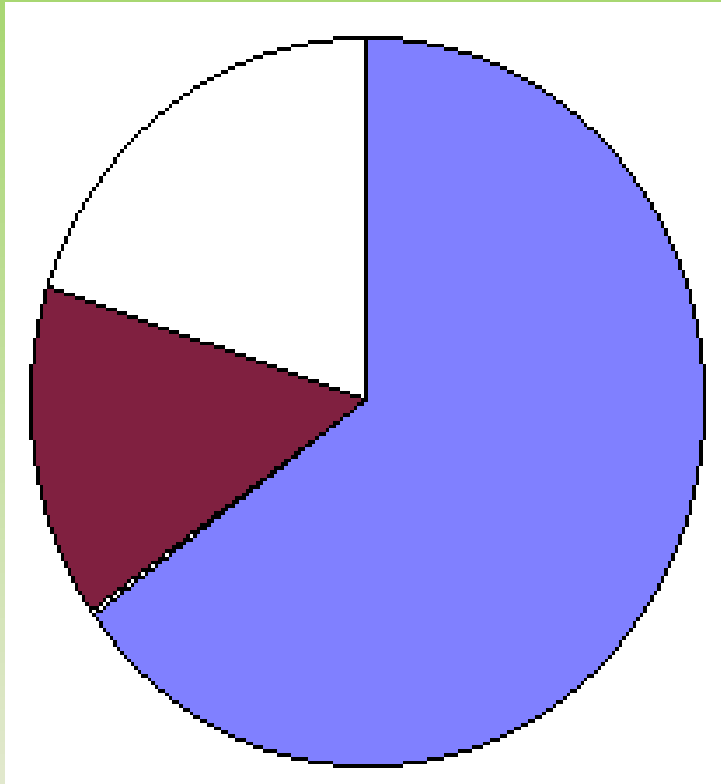
Amostragem Estratificada (AE)

- Amostragem estratificada uniforme



Amostragem Estratificada (AE)

- Amostragem estratificada proporcional



Amostragem em Estágios (AC)

- População considerada homogênea.
- Divisão em subgrupos semelhantes: os **conglomerados (estágios)**.
- Sorteiam-se os conglomerados:
 - Analisam-se todos os sorteados;
 - Sorteiam-se elementos dos conglomerados previamente sorteados.



Amostragem não probabilística

- Não há acesso a toda a população.
- Se as características da população acessível forem semelhantes às da população alvo: resultados “equivalentes” a uma amostragem probabilística.
- Utilizar apenas em casos particulares.

Exercício 1

Considerando a população de 50.000 colaboradores de uma empresa, formar uma amostra de 2.500 colaboradores para uma pesquisa de opinião.

Utilize a variável “cargo” para estratificar a população pela amostragem estratificada proporcional como critério para determinar o tamanho da amostra em cada estrato.

Cargo	No. colaboradores da empresa
A	5.000
B	15.000
C	30.000
Total	50.000

Cargo	No. colaboradores amostra
A	
B	
C	
Total	2.500

Exercício 2

Retirar uma amostra de 10 alunos e calcular a **altura** média utilizando as amostragens: **ASA**, **AS** e **AE**

ASA			AS			AE		
ID	altura		ID	altura		ID	altura	
2	1,70		2	1,70		2	1,92	
7	1,73		6	1,70		5	1,60	
10	1,75		10	1,75		10	1,70	
15	1,70		14	1,63		19	1,65	
18	1,74		18	1,74		21	1,70	
28	1,60		22	1,51		28	1,74	
30	1,56		26	1,69		37	1,89	
33	1,79		30	1,56		38	2,10	
43	1,76		34	1,78		41	1,60	
45	1,81		38	1,70		45	1,81	
média	1,71			1,68			1,77	
							grupo 1	
							grupo 2	

Dimensionamento da Amostra

- 1) **Tamanho da população** - pode-se classificá-la em finita e infinita.
- 2) **Variância ou percentual** - Variância (variáveis quantitativas) e percentual (variáveis qualitativas).
- 3) **Nível de confiança** - em geral, utiliza-se o nível de 95% de confiança, cujo valor é dado por $z = 1,96$.
- 4) **Margem de erro ou precisão** - à diferença entre o valor da estatística e o valor do parâmetro.
- 5) **Informação na literatura** - informações obtidas na literatura, podem dar indicativos dos valores a serem utilizados na determinação do valor de n .

Dimensionamento da Amostra

- Caso particular: Estimação de Prevalência

População N =	Margem de Erro	
	d=3 p.p. n =	d=5 p.p. n =
100	91	79
1.000	516	277
5.000	879	356
20.000	1013	377
100.000	1055	383
500.000	1064	384
1.000.000	1067	384

N = tamanho da população

n = tamanho da amostra

p.p. pontos percentuais

Dimensionamento da Amostra

Estimação de Prevalência: Variável **Qualitativa**

$$n_0 = \frac{z^2 p (1 - p)}{d^2}$$

n_0 = tamanho inicial da amostra

z = nível de confiança
($z = 1,96$ para 95% de confiança)

p = probabilidade ou prevalência (proporção)

d = margem de erro em p.p.

N = tamanho da população

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

Exercício 3

Determinar o tamanho da amostra para estimar a proporção de estudantes de P&E que **possuem notebook**

$$n_0 = \frac{z^2 p (1 - p)}{d^2}$$

$$\begin{cases} z = \\ p = \\ d = \end{cases}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

N	45	Tamanho da Amostra - Var. qualitativa	
p	0,5	n0= 1067	n= 43
z	1,96		
d(p)	0,03		

0,03 = 3% de margem de erro

Dimensionamento da Amostra

Estimação de Média: Variável **Quantitativa**

$$n_0 = \frac{z^2 \sigma^2}{d^2}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

n_0 = tamanho inicial da amostra

σ = desvio padrão populacional

z = nível de confiança

($z = 1,96$ para 95% de confiança)

d = margem de erro (na mesma unidade da variável)

N = tamanho da população

σ = medida de variabilidade

Exercício 4

Determinar o tamanho da amostra para estimar a **altura** média dos alunos dessa turma, com margem de erro de 0,06 m e nível de confiança de 95%, considere o desvio padrão igual a 0,11 m.

$$n_0 = \frac{z^2 \sigma^2}{d^2}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

$$\begin{cases} z = \\ \sigma = \\ d = \end{cases}$$

N	45	Tamanho da Amostra - Var. quantitativa			
σ	0,11	n0=	13	n=	10
z	1,96				
d(σ)	0,06				

Altura

Exercício 5

Determinar o tamanho da amostra para estimar a **idade** média dos alunos dessa turma, com margem de erro de 2 anos e nível de confiança de 95%, considere o desvio padrão igual a 6 anos.

$$n_0 = \frac{z^2 \sigma^2}{d^2}$$

$$n = \frac{n_0}{1 + \frac{n_0}{N}}$$

$$\begin{cases} z = \\ \sigma = \\ d = \end{cases}$$

N	45	Tamanho da Amostra - Var. quantitativa			
σ	6	n0=	35	n=	20
z	1,96				
d(σ)	2				
					Idade

Exercício 6

Com o tamanho da amostra calculado no exercício 5, faça o sorteio usando o método ASA, estime a **idade** média dos alunos desta sala de aula e construa o intervalo de precisão com a margem de erro de 2 anos. Utilize o arquivo *base_do_questionario_2025.xlsx*.

Exercício 6

ID_SORTEADO	IDADE
1	26
5	20
7	21
8	21
9	45
9	45
11	19
11	19
12	24
16	22
18	20
21	20
22	20
25	20
27	22
28	20
32	22
37	22
41	20
42	23
média \bar{x}	23,55
dp(s)	7,5
erro (d)	2

IC(95%) =	21,55	25,55
-----------	--------------	--------------

Intervalo de 95% de confiança que a verdadeira média da idade está no intervalo (21,55 ; 25,55) anos

O intervalo foi criado a partir da média amostral \bar{x} subtraindo e somando a margem de erro.

→ $\bar{x}-d; \bar{x}+d$

Exercício 7

Qual foi a margem de erro do exercício 4 ao sortear 10 alunos para calcular a **altura** média dos alunos da sala de aula? Considere o desvio padrão igual a 0,11 m.

$$\begin{cases} z = 1,96 \\ \sigma = 0,11 \\ n_0 = 13 \\ n = 10 \\ d = ? \end{cases}$$

$$n_0 = \frac{z^2 \sigma^2}{d^2}$$

Como analisar os dados de sua pesquisa?



	A	B	C	D	E	F	G	H	I	J	K
1	ID	CURSO	PESO	ALTURA	IDADE	NR_IRMAOS	CIDADE	TRABALHA	NOTEBOOK	GOSTA_ESTADISTICA	SATISFACAO
2	1	BCC	70	1,73	26	1	Blumenau	Não	Sim	Não sei	satisfeito/a
3	2	BEE	90	1,7	20	4	Blumenau	Não	Sim	Não sei	neutro/a
4	3	BEE	125	1,7	23	1	Blumenau	Sim	Sim	Não sei	satisfeito/a
5	4	BCC	89,1	1,71	20	0	Blumenau	Não	Sim	Sim	satisfeito/a
6	5	BEE	95	1,9	20	0	Blumenau	Sim	Sim	Sim	muito satisfeito/a
7	6	BEE	68	1,7	23	0	Blumenau	Sim	Sim	Sim	satisfeito/a
8	7	BCC	68	1,73	21	1	Blumenau	Não	Sim	Não sei	insatisfeito/a