

Estatística e Informática

Aula 02 - Introdução à Estatística

Alan Rodrigo Panosso alan.panosso@unesp.br

Departamento de Ciências Exatas FCAV/UNESP

07-03-2024

INTRODUÇÃO

Modelos Matemáticos

Modelos Determinísticos:

Nesse modelo as condições sob as quais um experimento é executado determinam o resultado do experimento.

Exemplo: Segunda Lei de Newton:

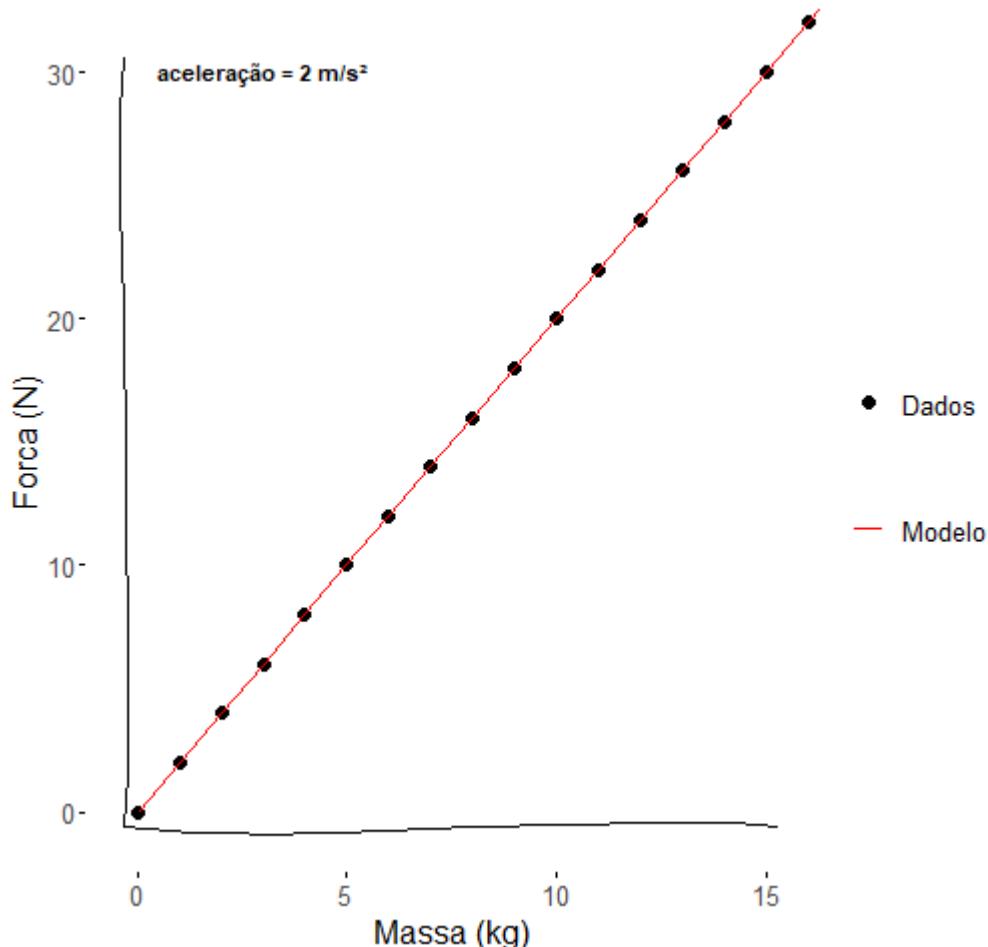
$$F = m \cdot a$$

Esse modelo diz que o valor da força F pode ser calculado tão logo os valores da massa do objeto m e sua aceleração a sejam fornecidos. Nesse tipo de modelo, quaisquer desvios que pudessem ocorrer seriam tão pequenos que a descrição acima seria suficiente para modelá-lo.

Seu resultado é determinado pelas condições sob as quais o experimento é executado.

Exemplo de mecânica clássica, astronomia, termodinâmica, circuitos elétricos e química são comuns e abundantes na literatura.

Modelo Deterministico



Dados = Modelo

Modelos Matemáticos

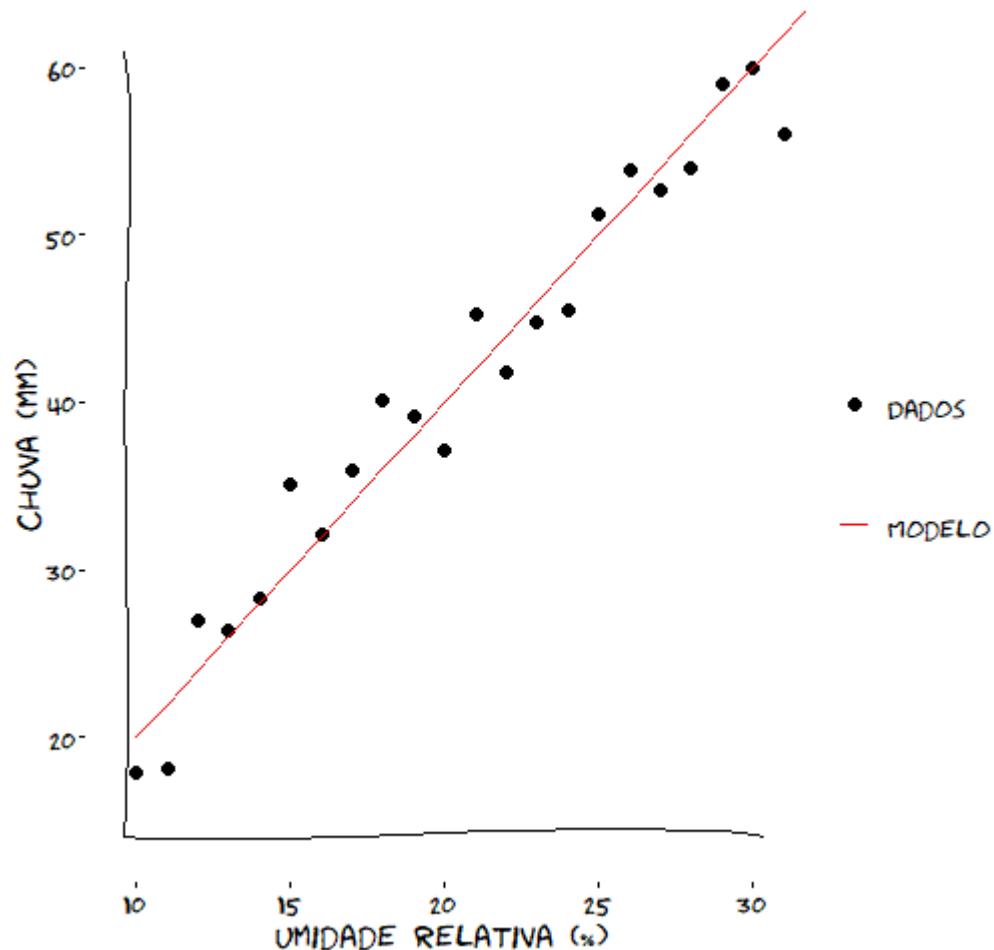
Para um grande número de situação na natureza, o modelo matemático determinístico apresentado é suficiente, contudo existem fenômenos que requerem um modelo matemático diferente:

Modelos Estocásticos (não-determinísticos ou probabilísticos):

Nesse modelo admitimos que as condições nas quais o ensaio é executado determinam somente o comportamento probabilístico do resultado observável.

Exemplo: em fenômenos meteorológicos, não podemos determinar qual será a precipitação em uma determinada região como resultado de uma tempestade. Observações como temperatura, pressão, velocidade do vento e umidade relativa do ar podem fornecer um prognóstico geral da chuva (fraca, média ou forte) entretanto, não tornam possível predizer quanta chuva cairá.

MODELO ESTOCASTICO (PROBABILISTICO)



$$Dados = Modelo + Ruído$$

A Estatística é fundamental na análise de dados provenientes de quaisquer processos onde exista variabilidade, estando interessada nos métodos e processos quantitativos que servem para a *coleta, organização, resumo, apresentação e análise desses dados*, bem como na **obtenção de conclusões válidas e na tomada de decisões** a partir de tais análises. Podendo ser dividida em três áreas:

Estatística Descritiva

Utilizada nas etapas iniciais dos trabalhos, se refere à maneira de **representar dados em tabelas e gráficos, resumi-los** por meio de algumas medidas **sem, contudo, tirar conclusões** sobre um grupo maior.

Probabilidade

Teoria matemática utilizada para se estudar a incerteza oriunda de fenômenos de caráter **aleatório**.

Inferência Estatística

Técnicas que possibilitam a extração das informações e conclusões obtidas a partir de subconjuntos de dados, a um grande número de dados. Ou seja, **conclusões aqui serão tomadas**.

ESTATÍSTICA DESCRIPTIVA

Estatística Descritiva

Utilizada nas etapas iniciais dos trabalhos, se refere à maneira de representar dados em tabelas e gráficos, resumi-los por meio de algumas medidas **sem, contudo, tirar conclusões sobre um grupo maior.**

É necessário, portanto, definirmos os nossos termos básicos para podermos nos comunicar durante a disciplina. Assim, serão definidos alguns conceitos.

Dados

Considerados o material básico da estatística, são os valores observados de uma característica de interesse de cada amostra, é o registro da característica de interesse.

Exemplos

- Variação de temperatura no processo de secagem de um alimento.
- Tempo de reabilitação de doentes após determinados tratamentos.
- Número de produtos defeituosos em lotes oriundos de uma linha de montagem.
- Peso e alturas de plantas de uma determinada variedade após aplicação de um trato cultural.

Exemplo de um banco de dados

Dados ← Clique para baixar o banco de dados no Excel
(exemplo_dados.xlsx)

A	B	C	D	E	F	
1	id	sexo	cor_cabelo	GA	altura	idade
2	1	F	CC	mais_social	1.68	19
3	2	F	CE	nao_consumes	1.59	20
4	3	F	CC	pouco	1.7	49
5	4	F	CE	socialmente	1.5	20
6	5	M	CE	mais_social	1.76	23
7	6	M	CC	pouco	1.6	28
8	7	M	L	nao_consumes	1.84	19

Onde:

id é a identificação do aluno;

sexo: "M" masculino e "F" feminino;

cor_cabelo: "P" - Preto, "C" - Castanho, "CC" - Castanho Claro, "CE" - Castanho Escuro, "L" - Loiro, "R" - Ruivo";

GA: é o grau de alcoolismo: "nao_consumes" < "pouco" < "socialmente" < "mais_social";

altura: altura declarada pelo aluno em *m*;

idade: idade do aluno ao cursar a disciplina (Estatística e Bioestatística).

Exemplo de um banco de dados

Dados ← Clique para baixar o banco de dados no R (**exemplo_dados.rds**)

```
# Lendo o Banco de dados
dados_turmas<-readr::read_rds("../data/dados_turmas.rds")

# Mostrando os 7 primeiros registros
head(dados_turmas, n=7)
```

```
## # A tibble: 7 × 6
##       id sexo cor_cabelo grau_alcool altura idade_anos
##   <dbl> <chr> <chr>        <chr>     <dbl>      <dbl>
## 1     1 F     CC          mais_social  1.68      19
## 2     2 F     CE          nao_consumo  1.59      20
## 3     3 F     CE          socialmente  1.5       20
## 4     4 M     CE          mais_social  1.76      23
## 5     5 M     CC          pouco        1.6       28
## 6     6 M     L           nao_consumo  1.84      19
## 7     7 M     CE          pouco        1.88      20
```

População

Populações Comuns: Uma população é um conjunto de pessoas (ou elementos) que possuem uma característica observável comum, geralmente, de interesse de estudo. Conjunto total de objetos que queremos estudar, para os quais uma decisão deve ser tomada.



População

Populações Estatísticas: a população estatística se refere a dados, e não às pessoas (elementos) nessa abordagem, a população é composta de características dos elementos, independente de terem sido medidas ou não.

- Valores de pressão sanguínea dos indivíduos de uma população.
- Presença ou ausência de doenças na população de cervos do pantanal.



População

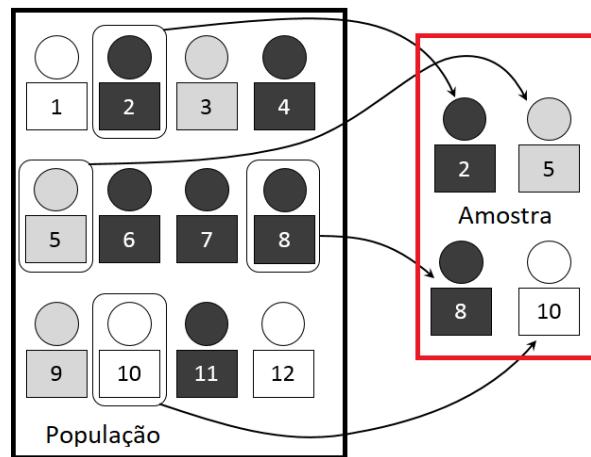
Populações Estatísticas:

- Porcentagem de controle de praga por inseto predador em uma parcela experimental.
- Valores de produção de soja por estados no Brasil.



Amostra

Amostra é qualquer subconjunto da população.



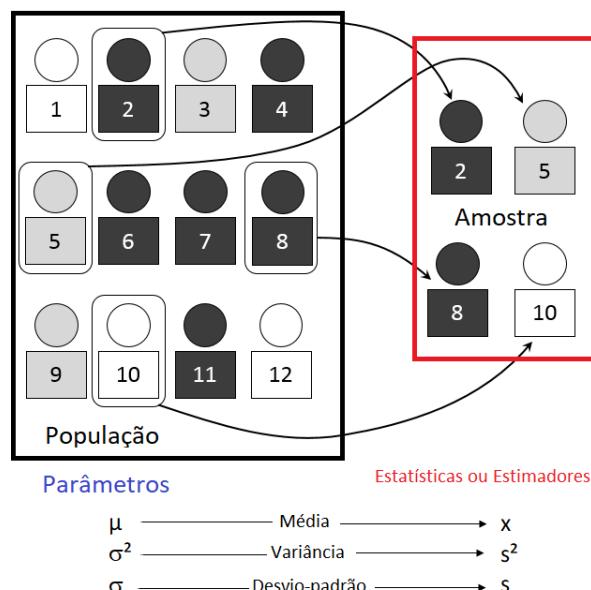
- Não conseguimos acessar toda uma população para estudar as características de interesse, tomaremos alguns elementos dessa população para formar um grupo a ser estudado.
- Impossibilidade de acesso e implicações éticas.

Parâmetro:

É a medida usada para descrever uma característica da população, por exemplo: média populacional (μ) ou a variância populacional (σ^2).

Estatística

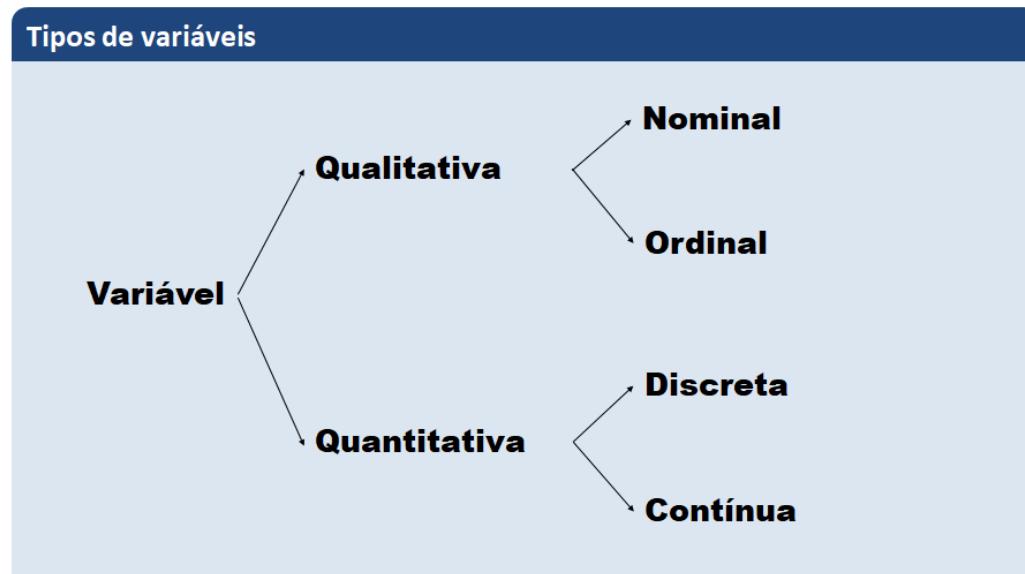
É a medida usada para descrever uma característica da amostra, em analogia, média amostral (\bar{x} ou \bar{m}) e a variância amostral (s^2).



Variável

As informações obtidas, sejam com base nos elementos que constituem a população, sejam com base nos elementos que constituem a amostra, são denominadas **dados**.

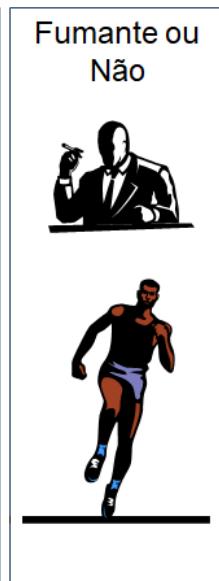
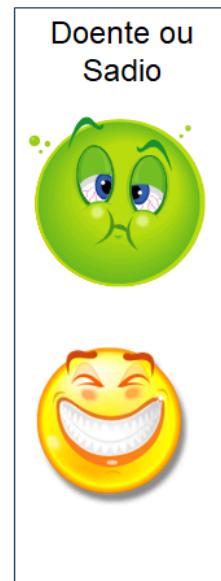
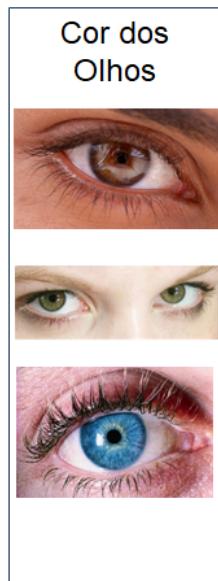
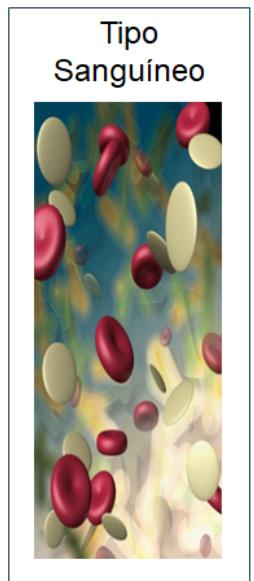
Assim, definimos que todo dado coletado refere-se a uma característica da população, agora, para nós determinada **VARIÁVEL**.



Variável Qualitativa

É aquela que apresenta como possíveis realizações uma qualidade (ou atributo) do indivíduo pesquisado.

Nominal: é aquela para a qual **não existe ordenação** alguma das possíveis realizações (característica observados). Em estatística dizemos que os dados são **categóricos**.



Variável Qualitativa

Ordinal: é aquela para a qual **existe certa ordem** nos possíveis resultados. Apesar de ordenar, não permite a indicação em termos de quanto mais ou menos.

Estádios de desenvolvimento de um inseto



Grau de Instrução



Níveis de infestação de doenças

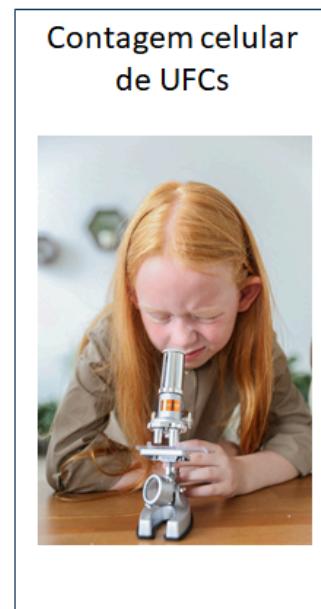
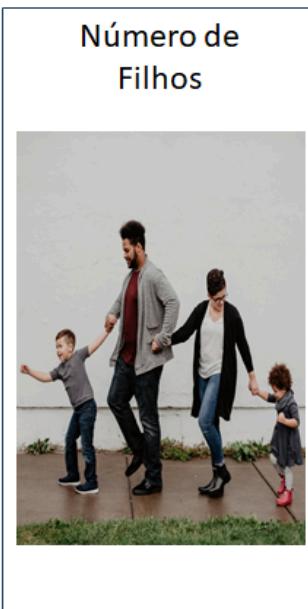
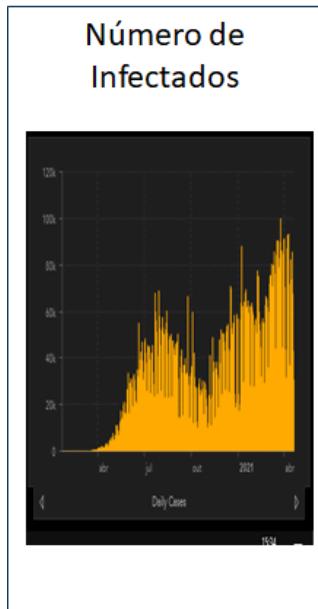


Datas (anos, mês e dia)



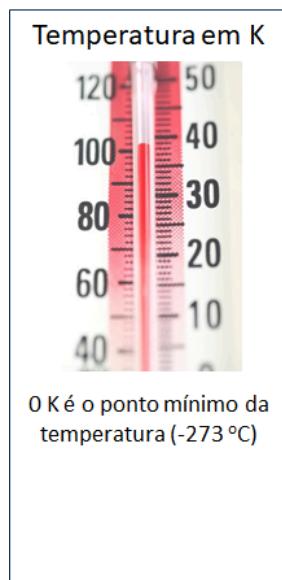
Variável Quantitativa

Discreta: os possíveis valores formam um conjunto enumerável resultam, frequentemente, de um **processo contagem** (os valores podem ser finito ou mesmo infinitos). Exemplos: número de filhos, número de células, número de ovos, número de ácaros ou insetos em uma planta.



Variável Quantitativa

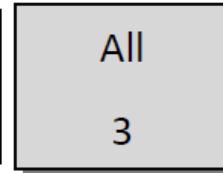
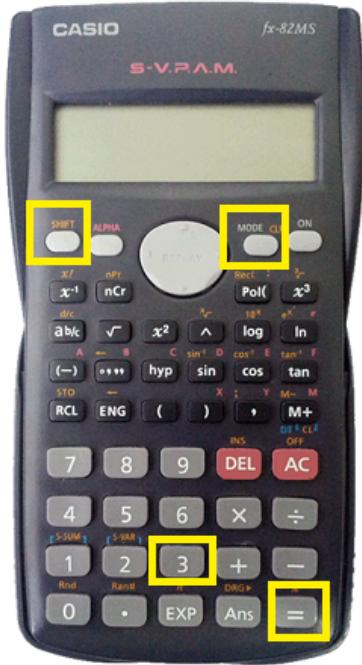
Contínua: os possíveis valores formam um intervalo de números reais e que resultam, normalmente, de um processo de medida (mensuração). Exemplos: peso, altura, produção de leite, pressão arterial, teor de nitrogênio no solo ou na folha.



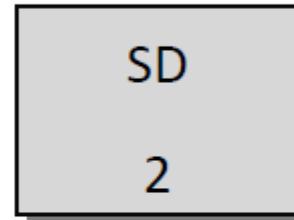
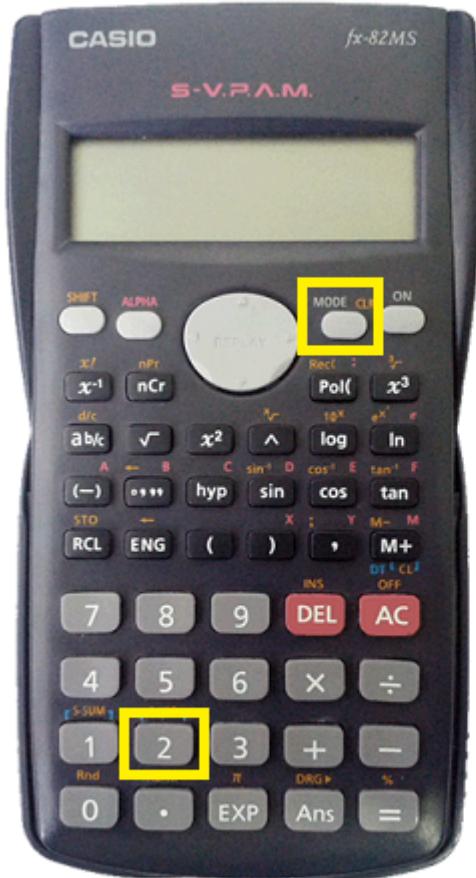
Ferramental

Calculadora Científica

Limpeza da Memória da Calculadora

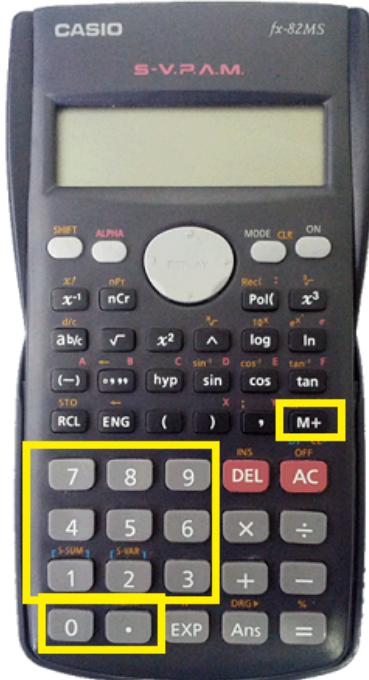


Acesso do modo Estatístico (SD)



Exemplo de entrada de dados

Entrar com os valores 3.25, 1.25, 3.20 e 2.30



3.25

1.25

3.20

2.30

M+

M+

M+

M+

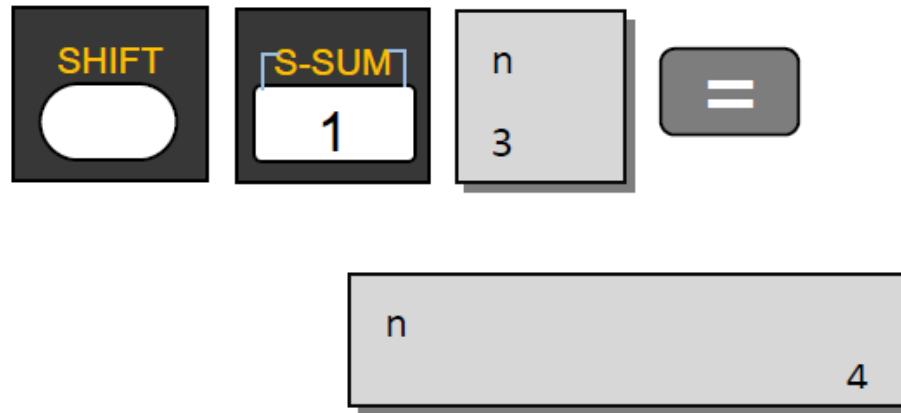
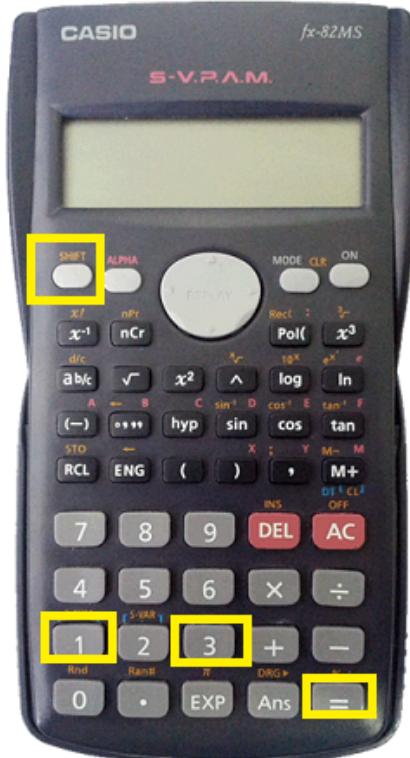
n= 1

n= 2

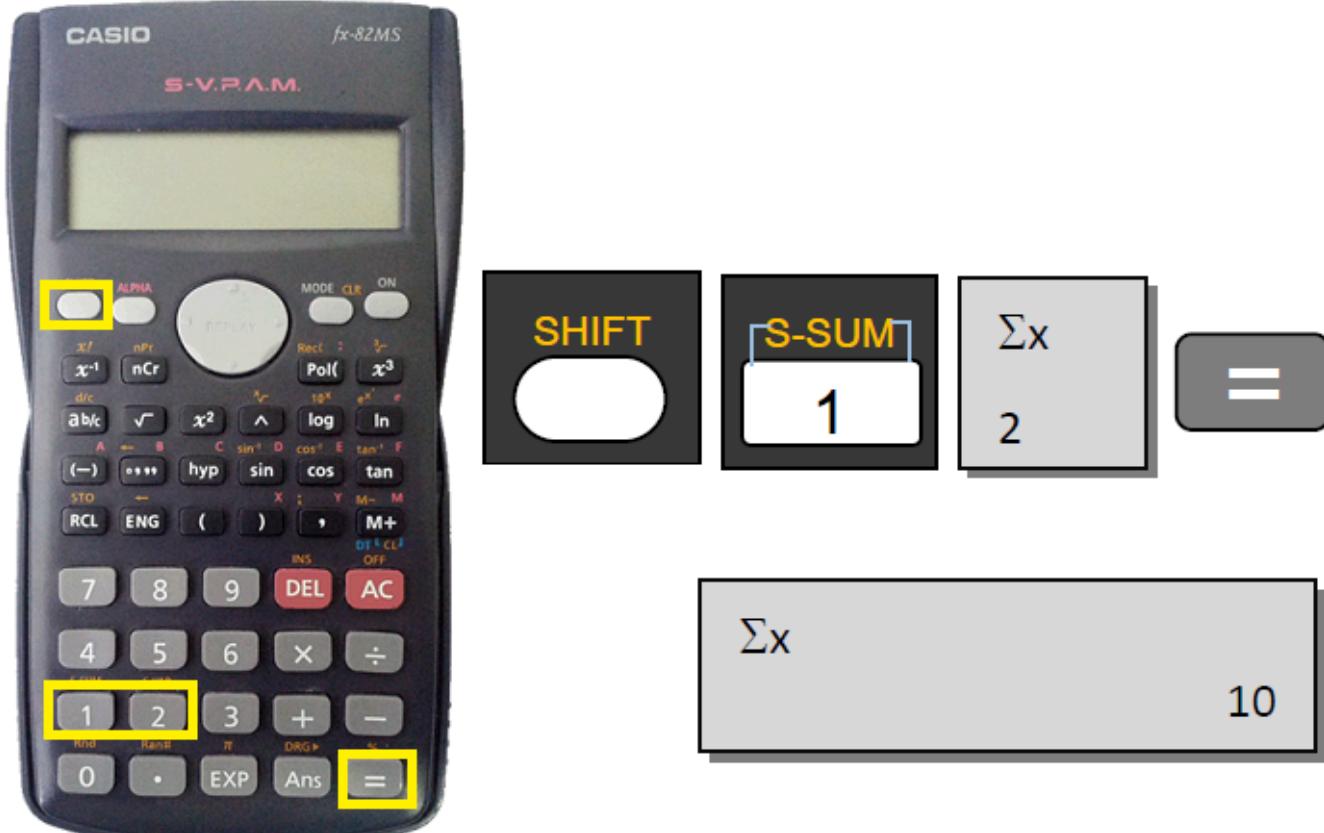
n= 3

n= 4

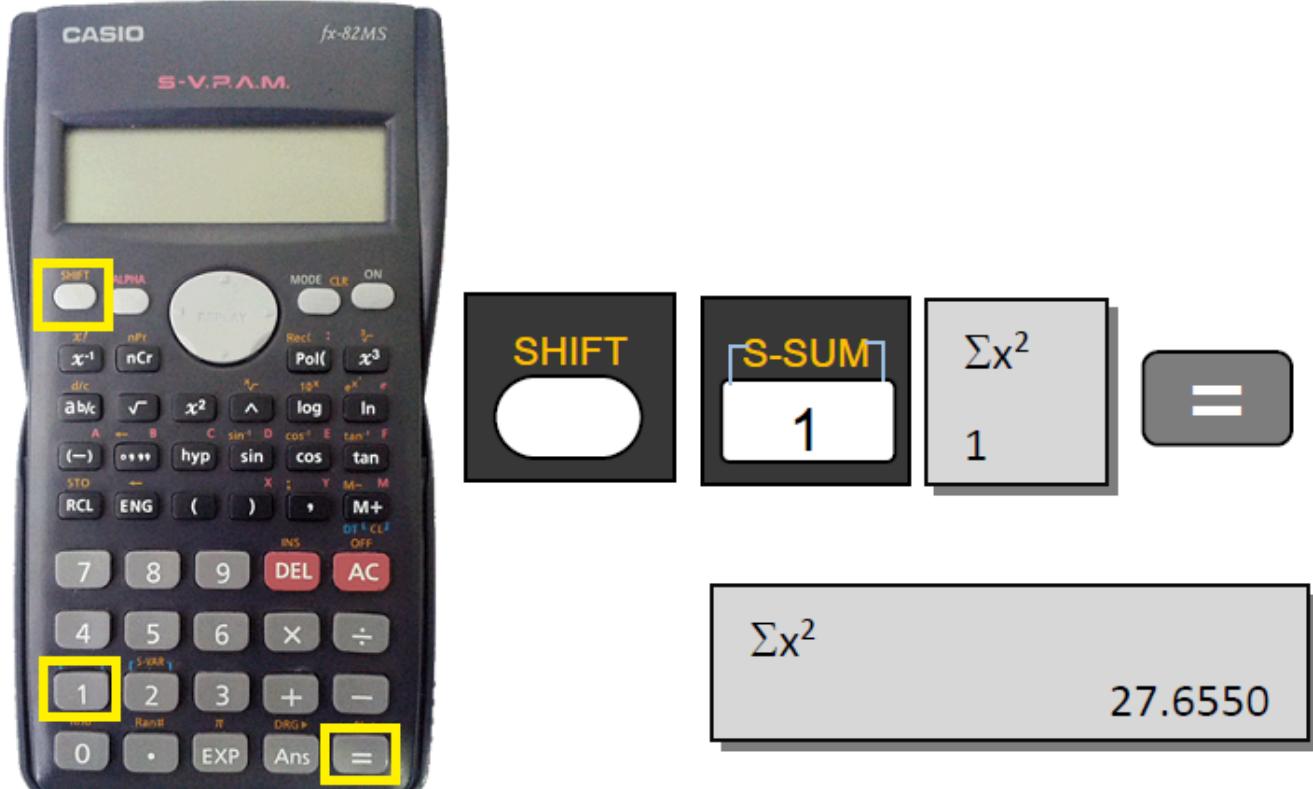
Número total de elementos na memória



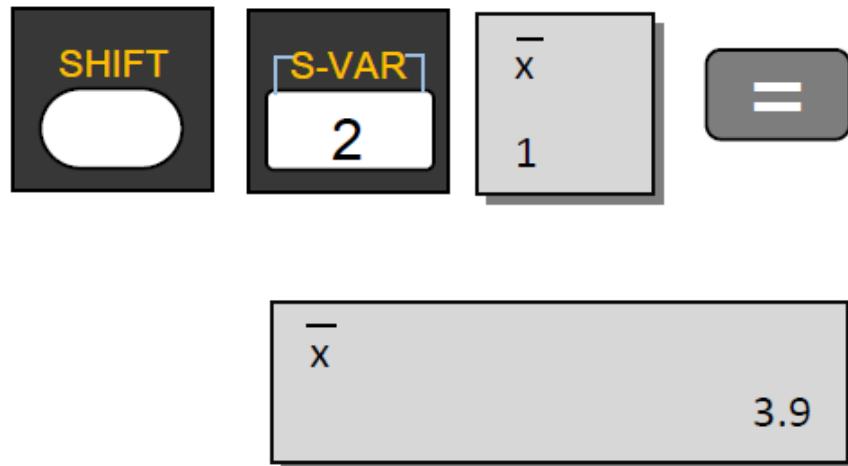
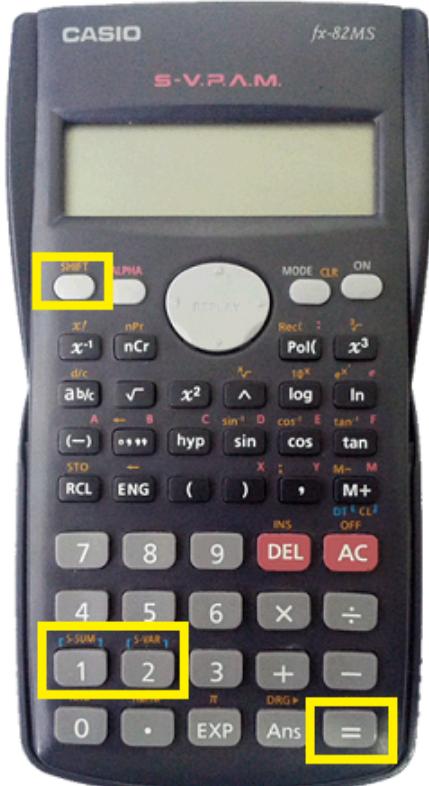
A somatória de todos os elementos ($\sum x$)



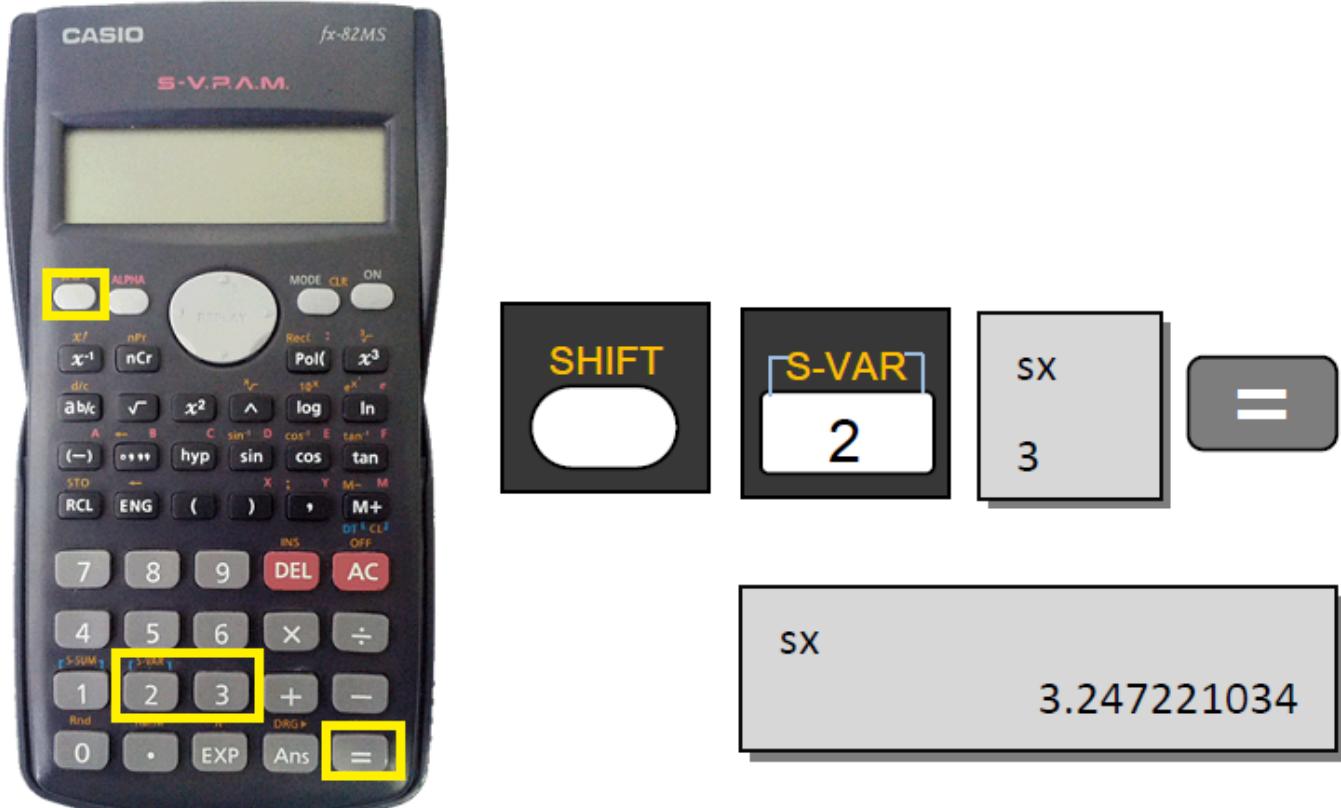
A somatória do quadrado de todos os elementos ($\sum x^2$)



Média dos valores da memória (\bar{x})



Desvio padrão amostral de todos os elementos (s) ou σ_x



1) Utilizando a Calculadora Científica e o R, monte as expressões para resolver as equações:

a) $\log_5 125$

b) $2 \cdot \sin(45^\circ)$

c) $\frac{1}{\sqrt{2\pi}}$

d) $\frac{1}{\sqrt{2\pi + 1}}$

2) Dados a amostra da variável $X = \{6, 8, 7\}$, sendo Σ a somatória dos elementos de X e \bar{x} a média (amostral) dos elementos de X e n o número total de elementos dessa amostra, calcule:

a) $Soma = \sum_{i=1}^n x_i$

b) $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

c) $\beta = \frac{(\sum_{i=1}^n x_i)^2}{(n - 1) \cdot (n - 2)}$

Resolução no R

Exercício 01

```
# 1.a  
log(125,5)
```

```
## [1] 3
```

```
# 1.b  
2*sin(45*pi/180)
```

```
## [1] 1.414214
```

```
# 1.c  
1/(sqrt(2*pi))
```

```
## [1] 0.3989423
```

```
# 1.d  
1/(sqrt(2*pi+1))
```

```
## [1] 0.3705436
```

Resolução no R

Exercício 02

```
# Definindo dados X  
X = c(6,8,7)
```

```
# 2.a  
sum(X)
```

```
## [1] 21
```

```
# 2.b  
mean(X)
```

```
## [1] 7
```

```
# 2.c  
sum(X)^2 / ((3-1)*(3-2))
```

```
## [1] 220.5
```