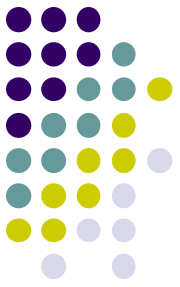


BIOESTATÍSTICA

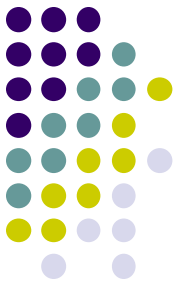
POPULAÇÃO E AMOSTRA



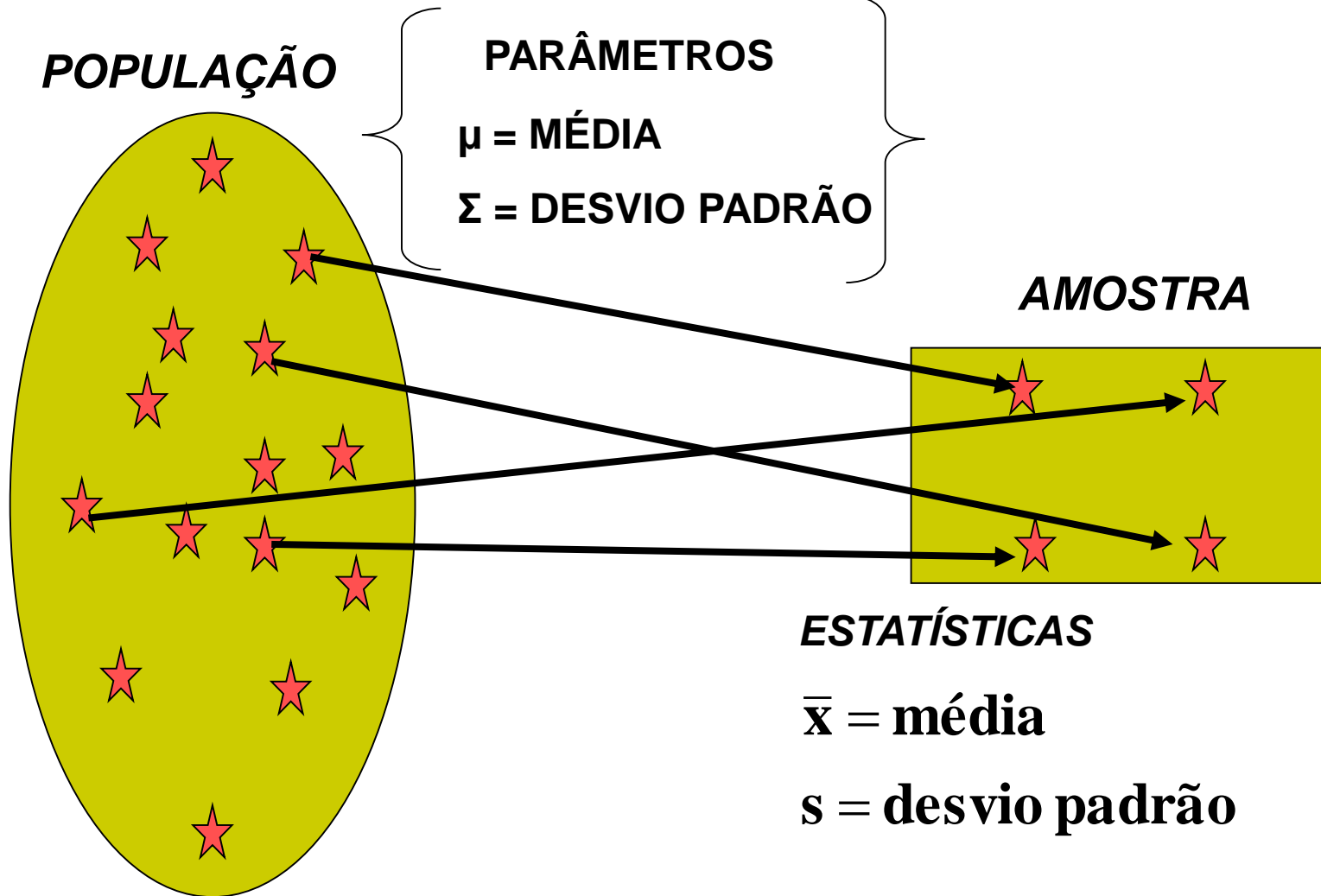
POPULAÇÃO: DEFINIÇÃO

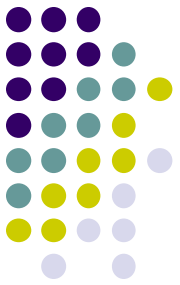


- **BIOLÓGICA:** todos os indivíduos de uma determinada espécie encontrados em uma área específica em um determinado período de tempo.
- **ESTATÍSTICA:** todos os possíveis valores de uma variável (também chamado Universo)
- **FINITA:** árvores em uma floresta, pessoas em uma cidade.
- **INFINITA:** experimentos



AMOSTRA (parte de uma população)





AMOSTRAGEM

- Fazemos inferência sobre a população
- Determinamos o grau de incerteza dessa influência.

ANTRAZ (OVELHAS)

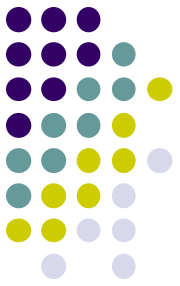
| | Doente | Sadia |
|--------------|--------|-------|
| Vacinado | 5 | 25 |
| Não Vacinado | 25 | 5 |

| | Doente | Sadia |
|--------------|--------|-------|
| Vacinado | 5 | 10 |
| Não Vacinado | 10 | 5 |

$$\chi^2 = 26,67, p < 0,0001$$

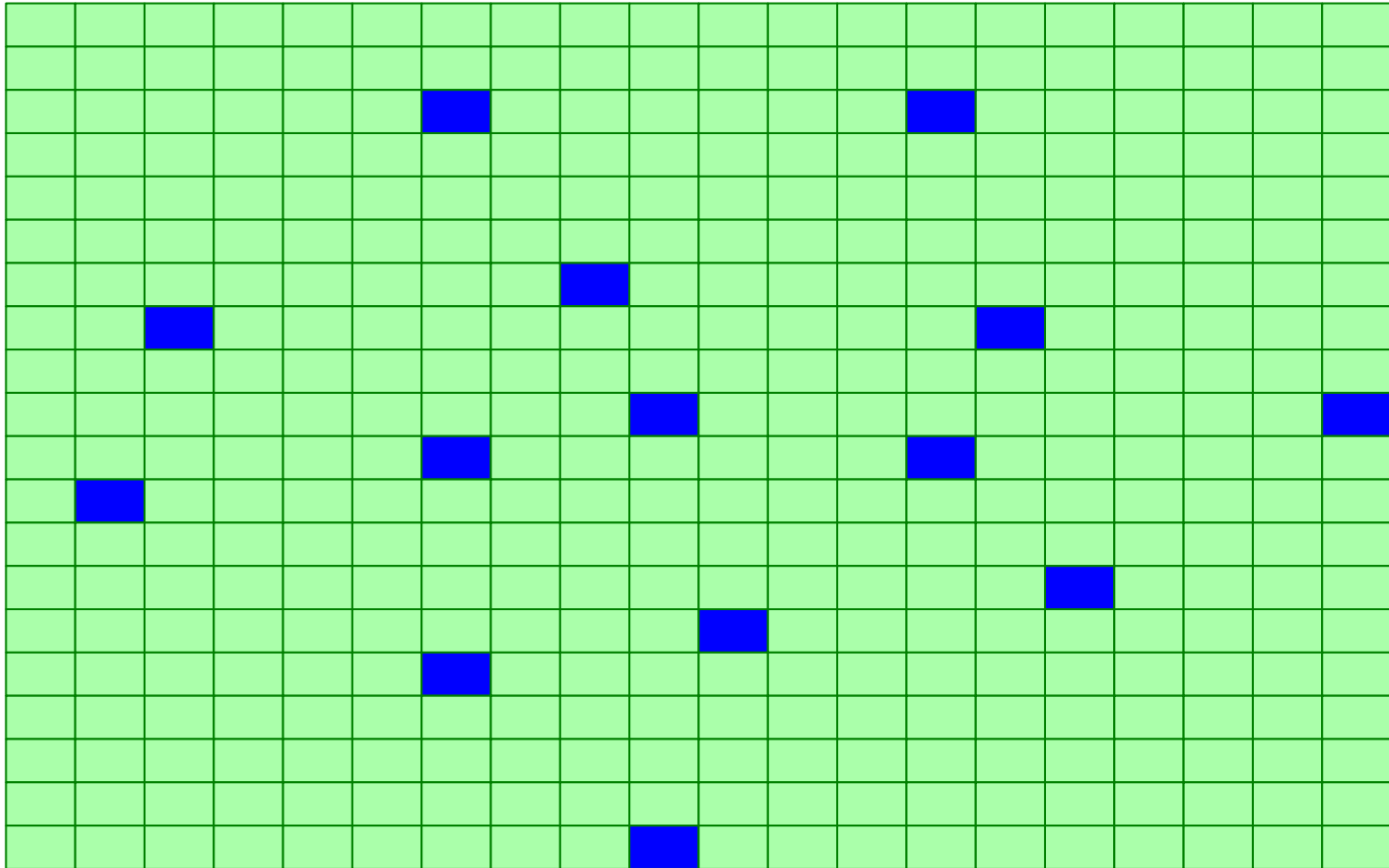
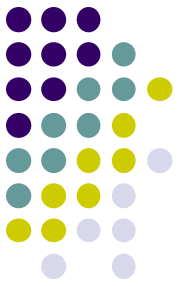
$$\chi^2 = 3,33, p = 0,0679$$

TIPOS MAIS COMUNS DE AMOSTRAGEM

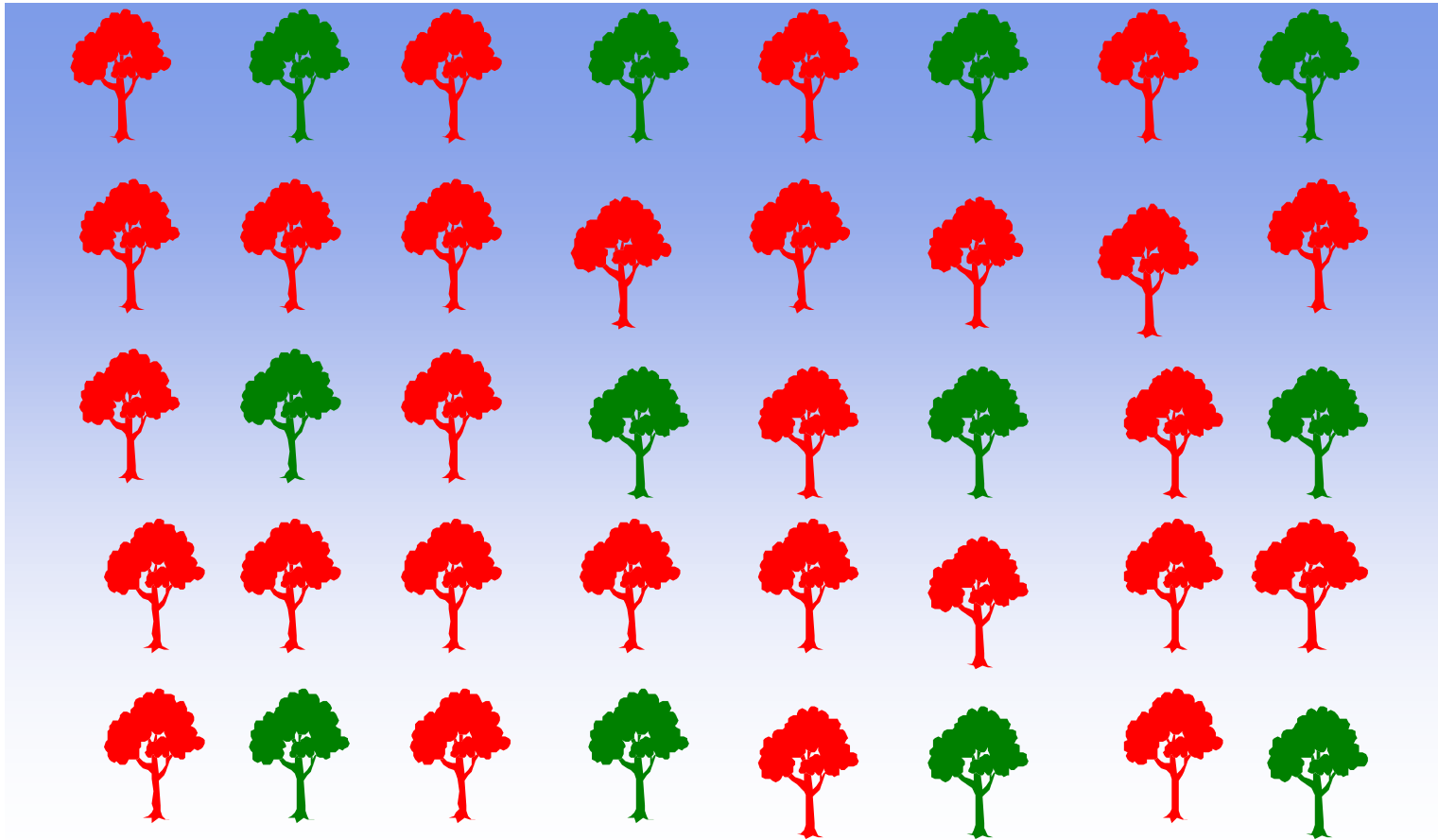
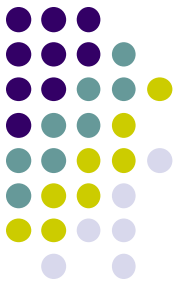


- ALEATÓRIA
- SISTEMÁTICA
- ESTRATIFICADA

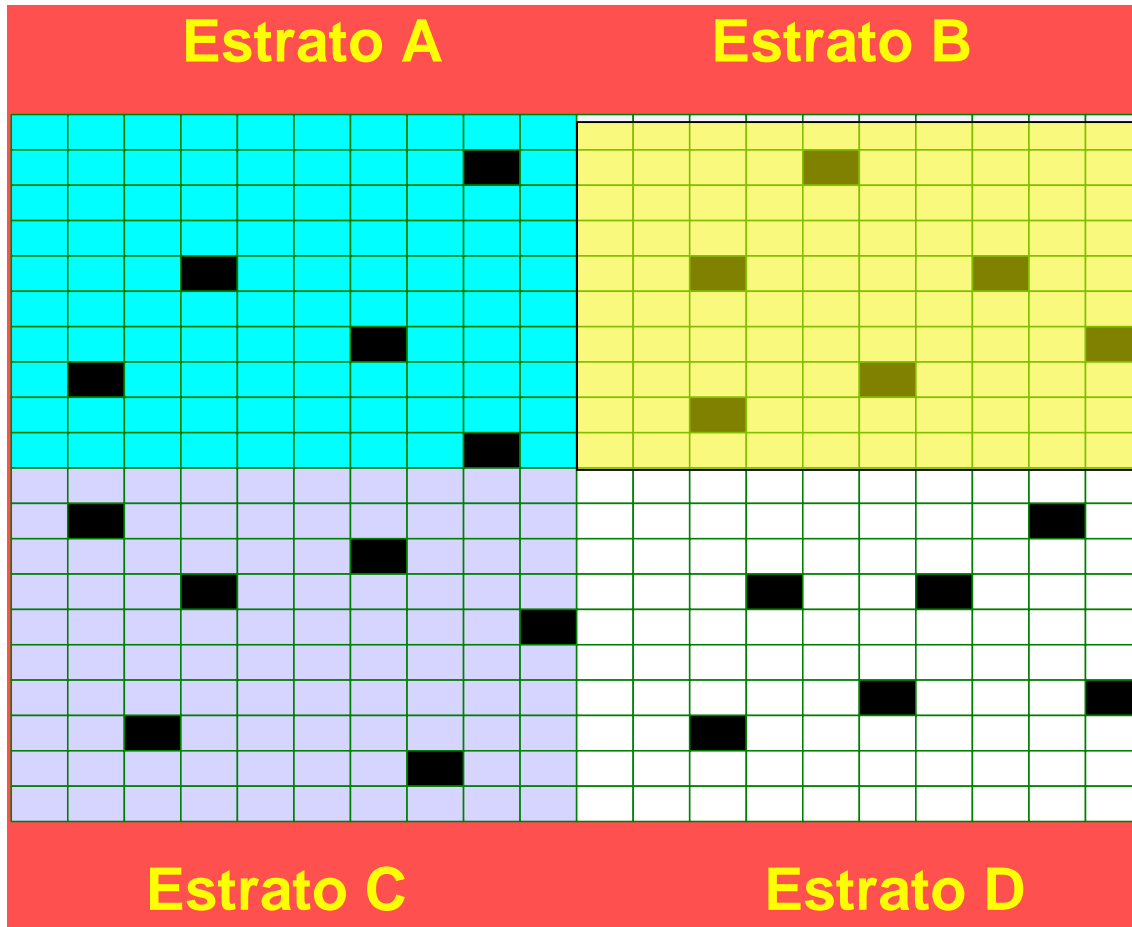
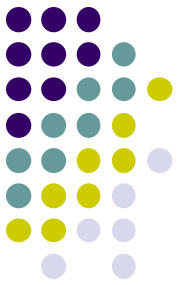
AMOSTRAGEM ALEATÓRIA



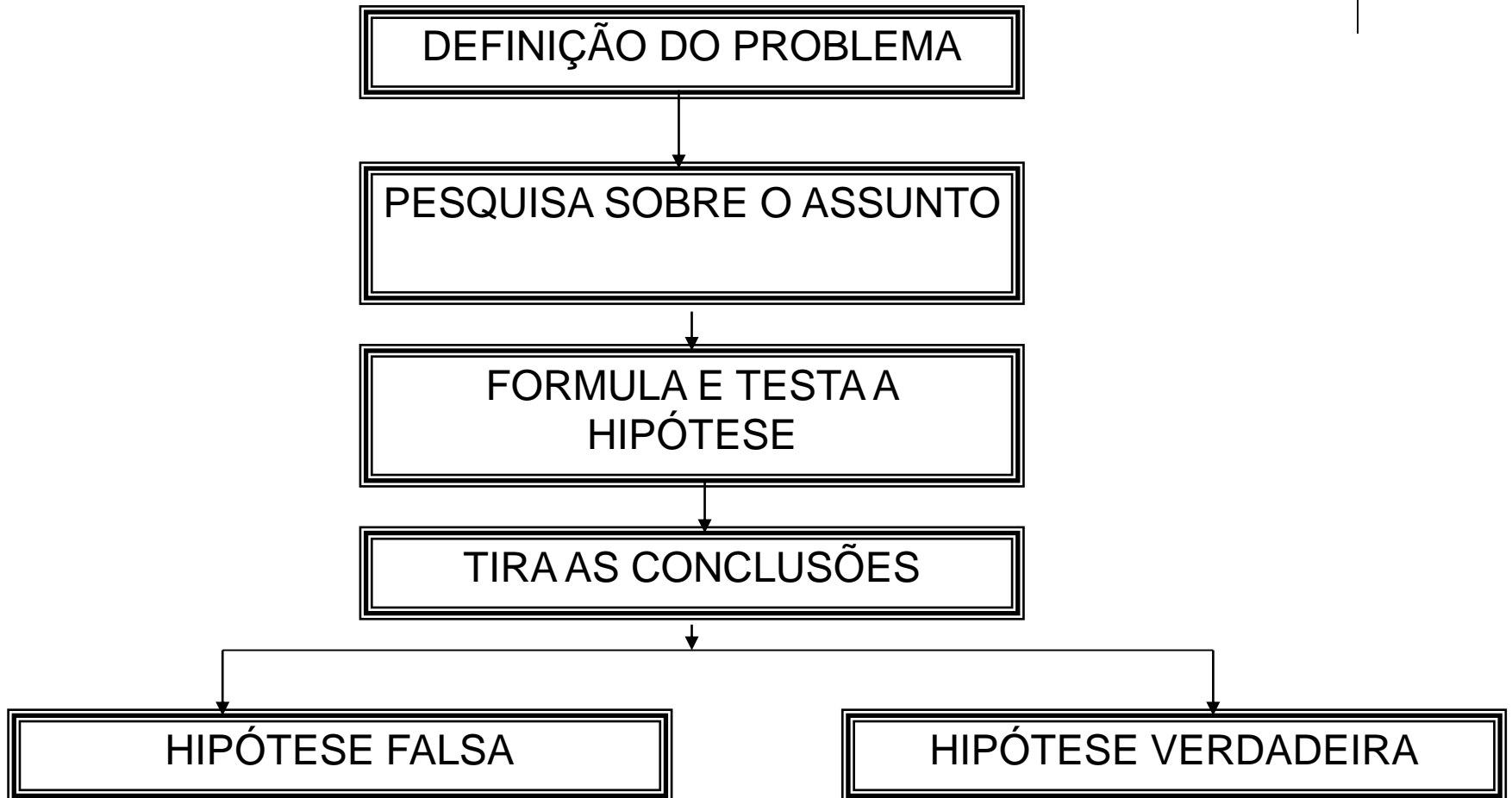
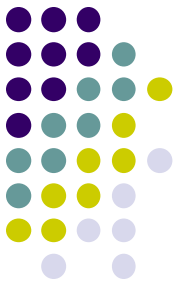
AMOSTRAGEM SISTEMÁTICA



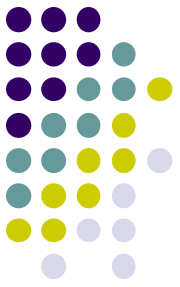
AMOSTRAGEM ESTRATIFICADA



MÉTODO CIENTÍFICO



PRECISÃO E EXATIDÃO



- EXATIDÃO (acurácia): proximidade de uma medida do seu valor real (viés)
- PRECISÃO: proximidade de medidas repetidas da mesma quantidade.



SEM VIÉS, MAS
IMPRECISA

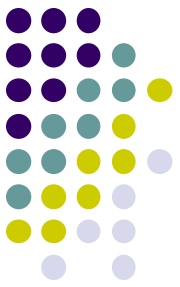


PRECISA, MAS COM
VIÉS



SEM VIÉS E
PRECISA

COMO DEVEMOS MEDIR UMA VARIÁVEL?



$$\bar{x} = 6,4538201 \text{ kg}$$

AMPLITUDE \rightarrow 4 – 8 KG \rightarrow 4 PASSOS UNITÁRIOS

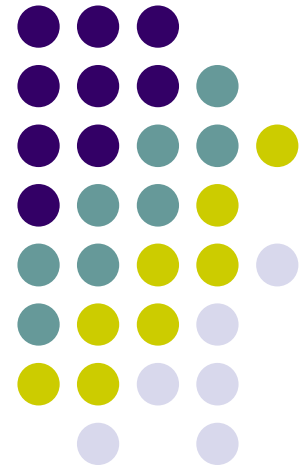
AMPLITUDE \rightarrow 4,1 – 8,2 \rightarrow 41 PASSOS UNITÁRIOS

RECOMENDA-SE DE 30 A 300 PASSOS UNITÁRIOS (DIFERENÇA ENTRE O MAIOR E MENOR VALOR MEDIDO).

Ex.: pH do solo varia de 3 a 6. Quantas casas decimais devemos medir?

BIOESTATÍSTICA

ESTADÍSTICA DESCRIPTIVA

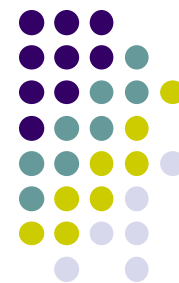


ESTATÍSTICA DESCRITIVA



- PARA A ANÁLISE EXPLORATÓRIA E DESCRIÇÃO DOS DADOS.
- PRINCIPAIS DESCRITORES: distribuição de frequência (método gráfico), e medidas de localização (tendência central) e dispersão (variabilidade).

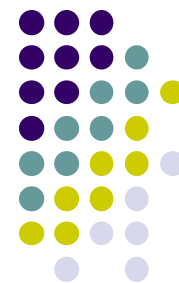
MÉTODO GRÁFICO



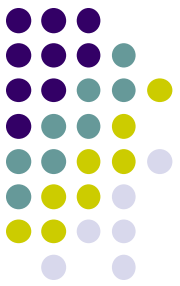
CONJUNTO DE DADOS (DAP DE ÁVORES NATIVAS) - cm

| | | | | |
|------|------|------|------|------|
| 20,5 | 19,5 | 15,6 | 24,1 | 9,9 |
| 15,4 | 12,7 | 5,4 | 17,0 | 28,6 |
| 16,9 | 7,8 | 23,3 | 11,8 | 18,4 |
| 13,4 | 14,3 | 19,2 | 9,2 | 16,8 |
| 8,8 | 22,1 | 20,8 | 12,6 | 15,9 |

MÉTODO GRÁFICO



- FREQUÊNCIA: absoluta (f)
acumulada (F) e relativa (f_n ou F_n)
- COMO DEVEMOS FAZER UM GRÁFICO DE FREQUÊNCIA?
- QUAL O NÚMERO DE CLASSES (k)?
- GERALMENTE ENTRE 5 E 20 CLASSES (MAIS COMUM ENTRE 7 E 13 CLASSES)
- $k = 1 + 3,319 \cdot \log_{10}(n)$ – Regra de Sturges para menos de 200 dados (JASA, 1926)
- $k = 1 + 3,319 \cdot \log_{10}(25) = 5,63 \approx 6$
- Para n maior que 200 usar a **Regra de Scott** (Biometrika, 1979):
- $k = 3,5 \cdot s \cdot n^{-1/3}$ (s = desvio padrão dos dados).
- No caso de dúvidas use a **Regra de Scott** (Hyndman, 1995)



MÉTODO GRÁFICO

$w = \textit{amplitude de classe}$

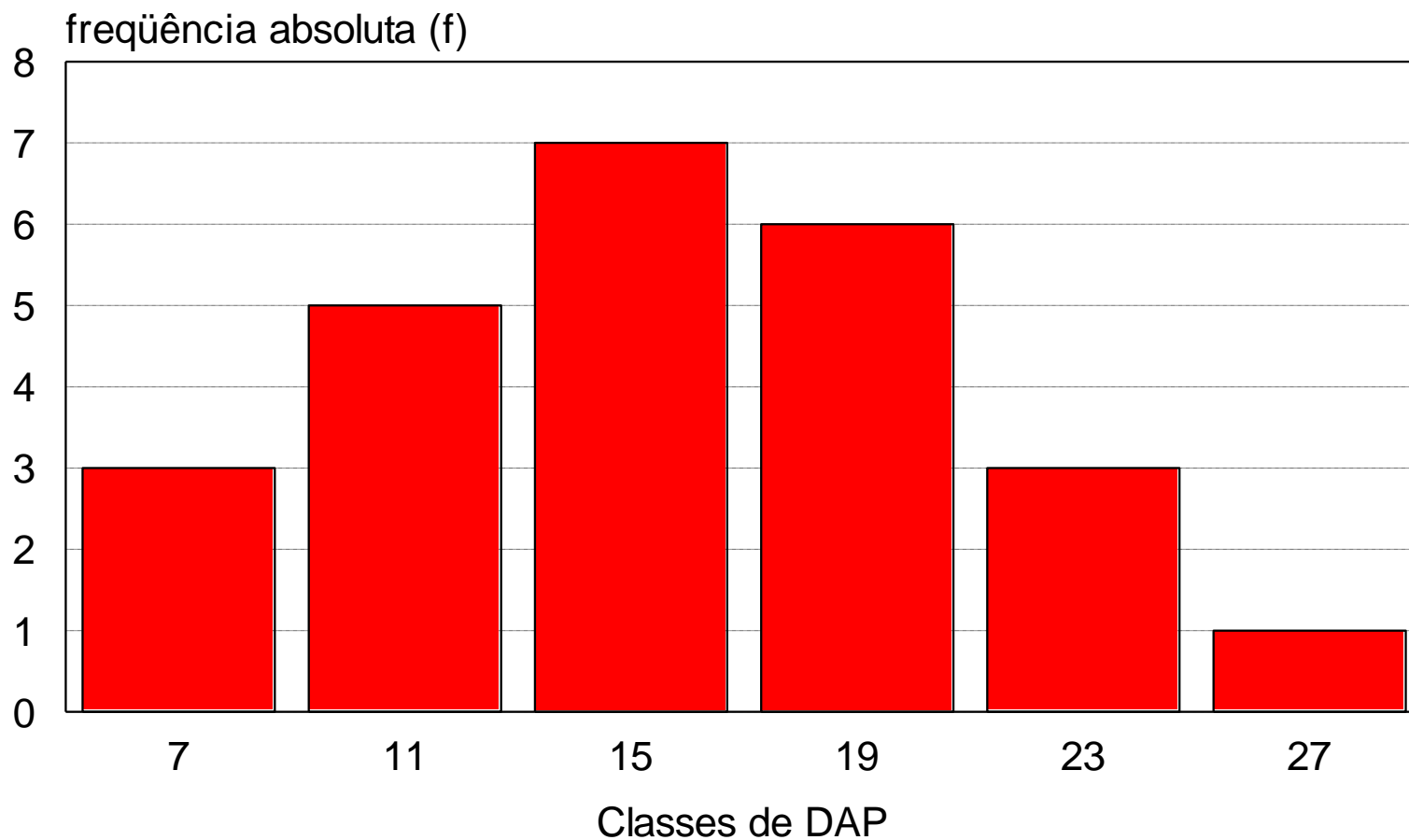
$$w = \frac{\max - \min}{k} = \frac{28,6 - 5,4}{6} = 3,87 \cong 4$$

| Classe | Amplitude | CC | f | f _n |
|--------|-----------|----|----|----------------|
| 1 | 5,0-8,9 | 7 | 3 | 0,12 |
| 2 | 9,0-12,9 | 11 | 5 | 0,20 |
| 3 | 13,0-16,9 | 15 | 7 | 0,28 |
| 4 | 17,0-20,9 | 19 | 6 | 0,24 |
| 5 | 21,0-24,9 | 23 | 3 | 0,12 |
| 6 | 25,0-28,9 | 27 | 1 | 0,04 |
| Total | | | 25 | 1,00 |

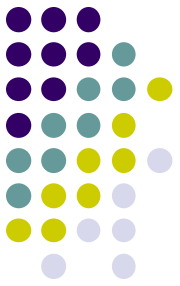
MÉTODO GRÁFICO



HISTOGRAMA DE FREQUÊNCIA



MEDIDAS DE LOCALIZAÇÃO



MÉDIA

CONJUNTO DE DADOS AMOSTRADOS: $x_1, x_2, x_3, \dots, x_n$

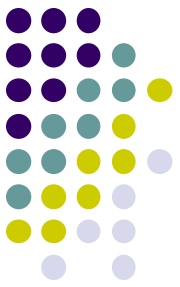
n = tamanho da amostra

x_i = valor da observação *i* na amostra

$$\text{Média} = \bar{x} = \frac{\mathbf{x}_1 + \mathbf{x}_2 + \mathbf{x}_3 + \dots + \mathbf{x}_n}{\mathbf{n}} = \frac{\sum \mathbf{x}_i}{\mathbf{n}}$$

$$\text{Propriedade : } \sum (\mathbf{x}_i - \bar{x}) = \mathbf{0}$$

MEDIDAS DE LOCALIZAÇÃO



MEDIANA

É O NÚMERO DO MEIO QUANDO AS OBSERVAÇÕES SÃO ARRANJADAS EM ORDEM CRESCENTE.

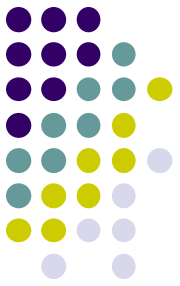
9, 2, 7, 11, 14 (n=5)

2, 7, 9, 11, 14 → med = 9

9, 2, 7, 11, 6, 14 (n=6)

2, 6, 7, 9, 11, 14 → med = 8 = (7 + 9)/2

MEDIDAS DE LOCALIZAÇÃO



MÉDIA PONDERADA

$$\bar{x}_w = \frac{\sum w_i \cdot x_i}{\sum w_i}$$

MP 100 ha

8 animais/ ha

MSI 60 ha

2 animais/ ha

4 animais/ ha

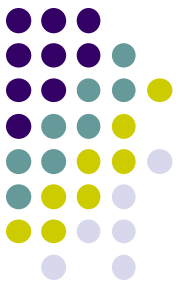
MST 50 ha

FRAGMENTO

MP = Mata primitiva

MST = Mata secundária tardia

MSI = Mata secundária inicial



MEDIDAS DE LOCALIZAÇÃO

MÉDIA PONDERADA

$$\bar{x}_w = \frac{\sum w_i \cdot x_i}{\sum w_i} = \frac{8 \cdot 100 + 2 \cdot 60 + 4 \cdot 50}{210}$$

$$\bar{x}_w = 5,3 \text{ animais/ha}$$

MP 100 ha

8 animais/ha

MSI 60 ha

2 animais/ha

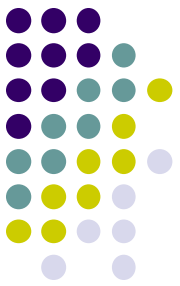
4 animais/ha

MST 50 ha

FRAGMENTO

Média não ponderada = 4,7 animais / ha

MEDIDAS DE DISPERSÃO



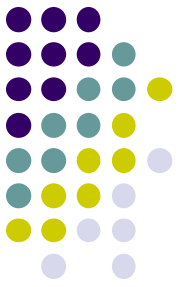
$$\text{VARIÂNCIA} = s^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$$

$$\text{DESVIO PADRÃO} = \sqrt{s^2} = s$$

$$\text{ERRO PADRÃO DA MÉDIA} = \sqrt{\frac{s^2}{n}} = s_{\bar{x}}$$

$$\text{COEFICIENTE DE VARIAÇÃO} = \frac{s}{\bar{x}} \cdot 100$$

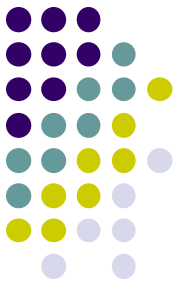
FÓRMULAS PARA DADOS AGRUPADOS:



$$\bar{x} = \frac{\sum x_i f}{n}$$

$$s^2 = \frac{\sum x_i^2 \cdot f - \frac{(\sum x_i \cdot f)^2}{n}}{n - 1}$$

TEOREMA DO LIMITE CENTRAL



Demonstrado por DeMoivre em 1733 para distribuição binomial (Gauss ainda não existia)

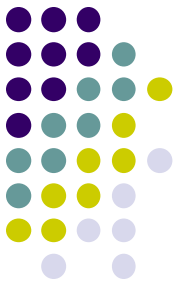
LaPlace em 1812 demonstrou que poderia ser generalizado para todas as demais distribuições

x_i = variável com qualquer distribuição

$$\bar{x} = \frac{\sum x_i}{n}$$

Possui distribuição normal, para amostras de tamanho grande (30 a 60 unidades amostrais)

TEOREMA DO LIMITE CENTRAL

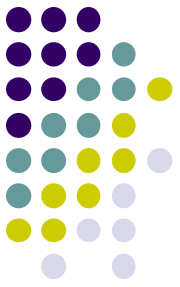


Abraham de Moivre

Nasceu na França em
1667 e faleceu na
Inglaterra em 1754

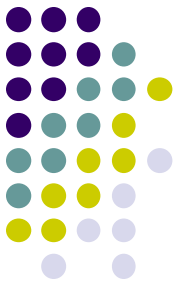
Pesquisador em
probabilidade, publicou
em 1718, na Inglaterra, o
livro *The Doctrine of
Chance*

Demonstração do Teorema do Limite Central

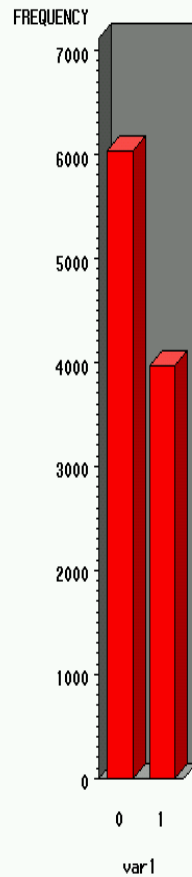


http://onlinestatbook.com/stat_sim/sampling_dist/

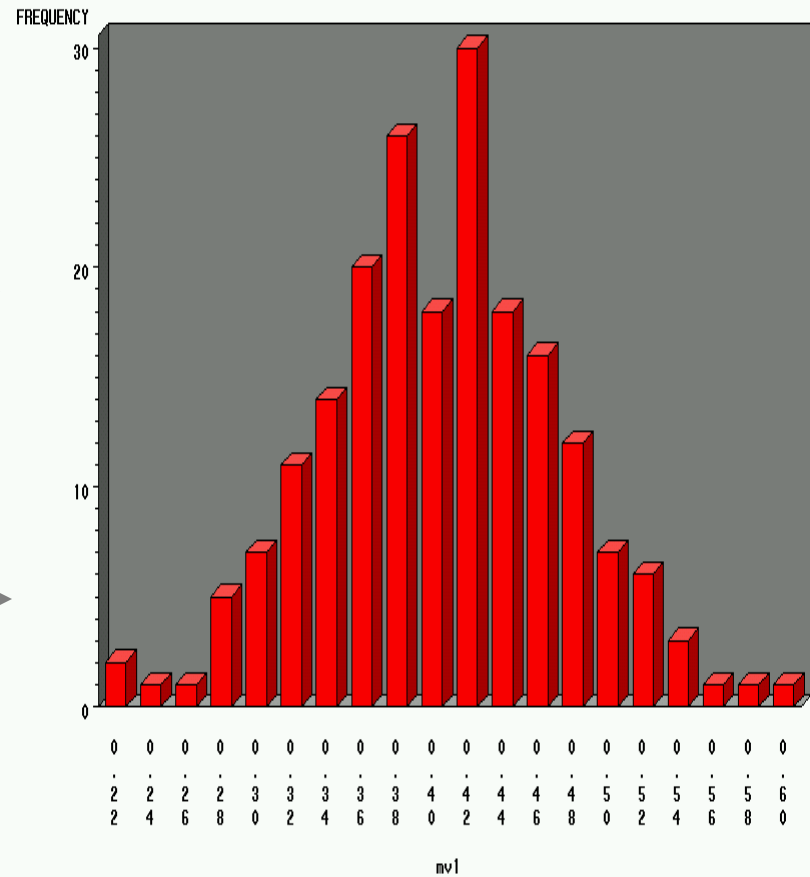
DISTRIBUIÇÃO BINOMIAL



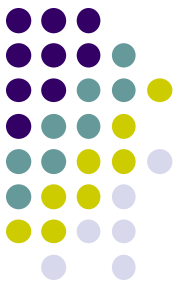
B
I
N
O
M
I
A
L



100
amostras
de
tamanho
50

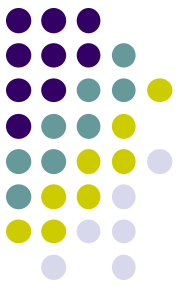


Qual o comportamento de um dado?



**Existe dado
honesto e
desonesto?**



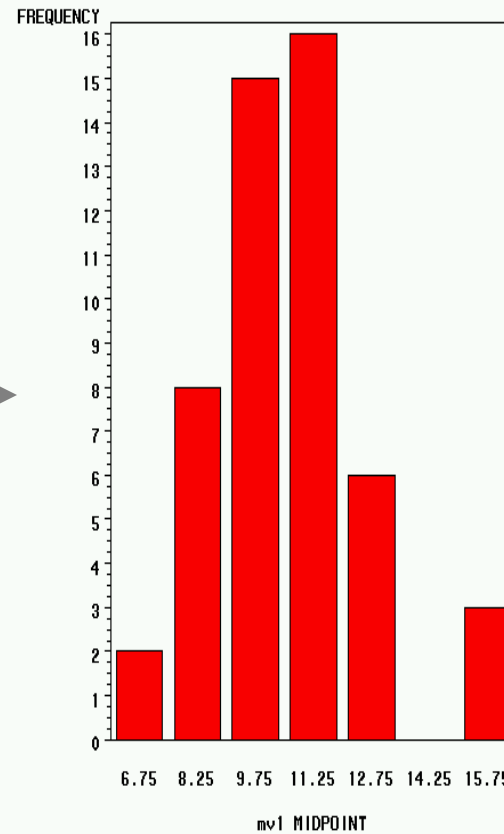
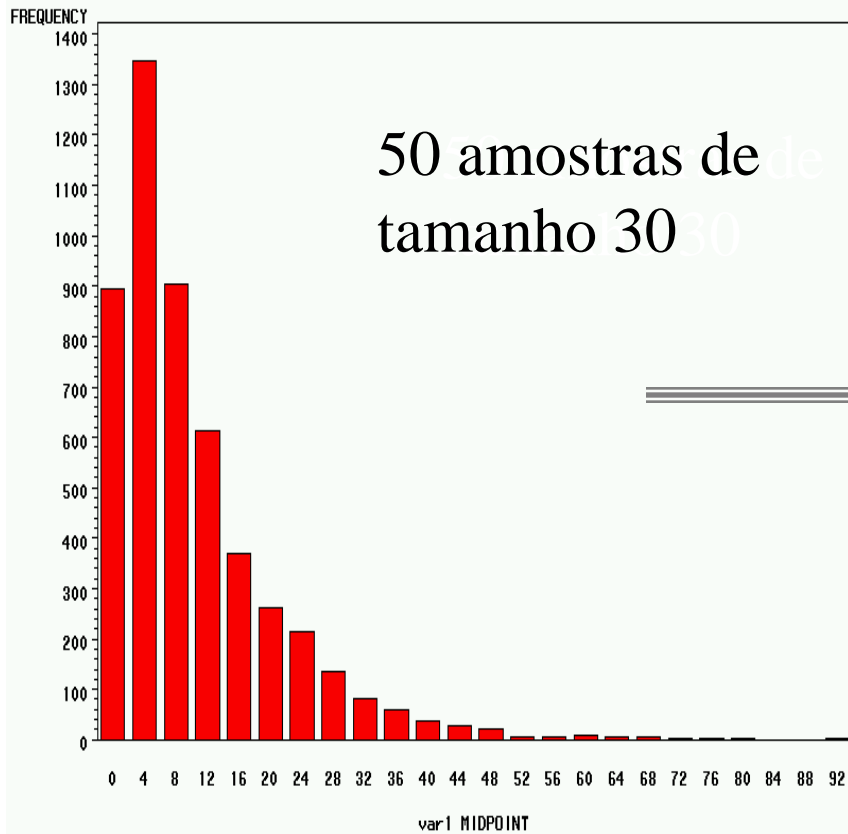
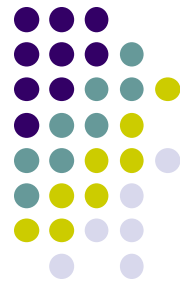


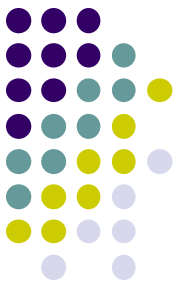
Pierre-Simon Laplace

Viveu na França de 1749 a 1827.

Pesquisava diferentes áreas desde equações diferenciais e probabilidade até mecânica celeste. Ainda trabalhou com Lavoisier no estudo comparativo do poder calorífico do arroz.

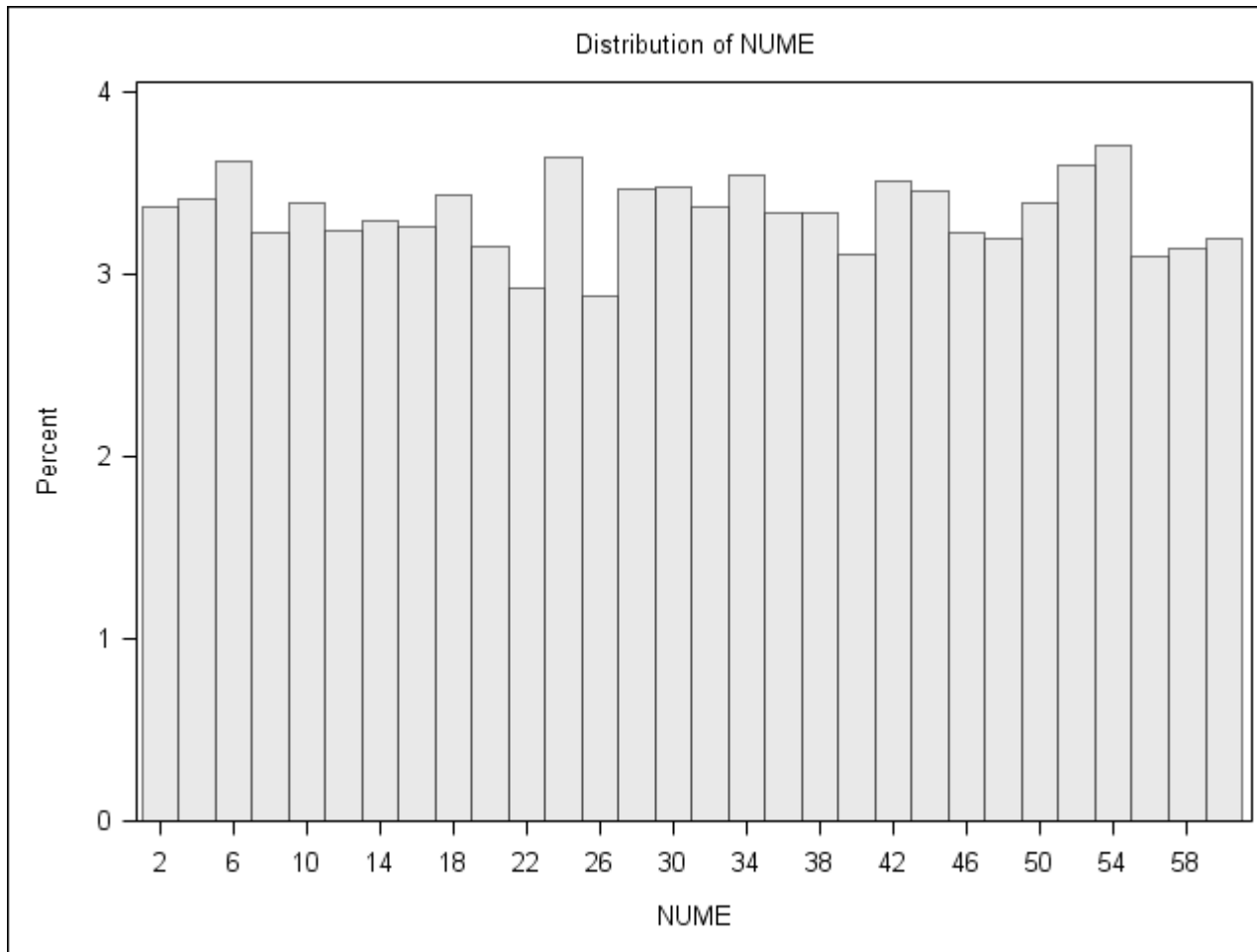
Distribuição Exponencial:





Megasena

Distribuição dos números sorteados.



29/01/2018

Megasena

Distribuição da média dos números sorteados por concurso.

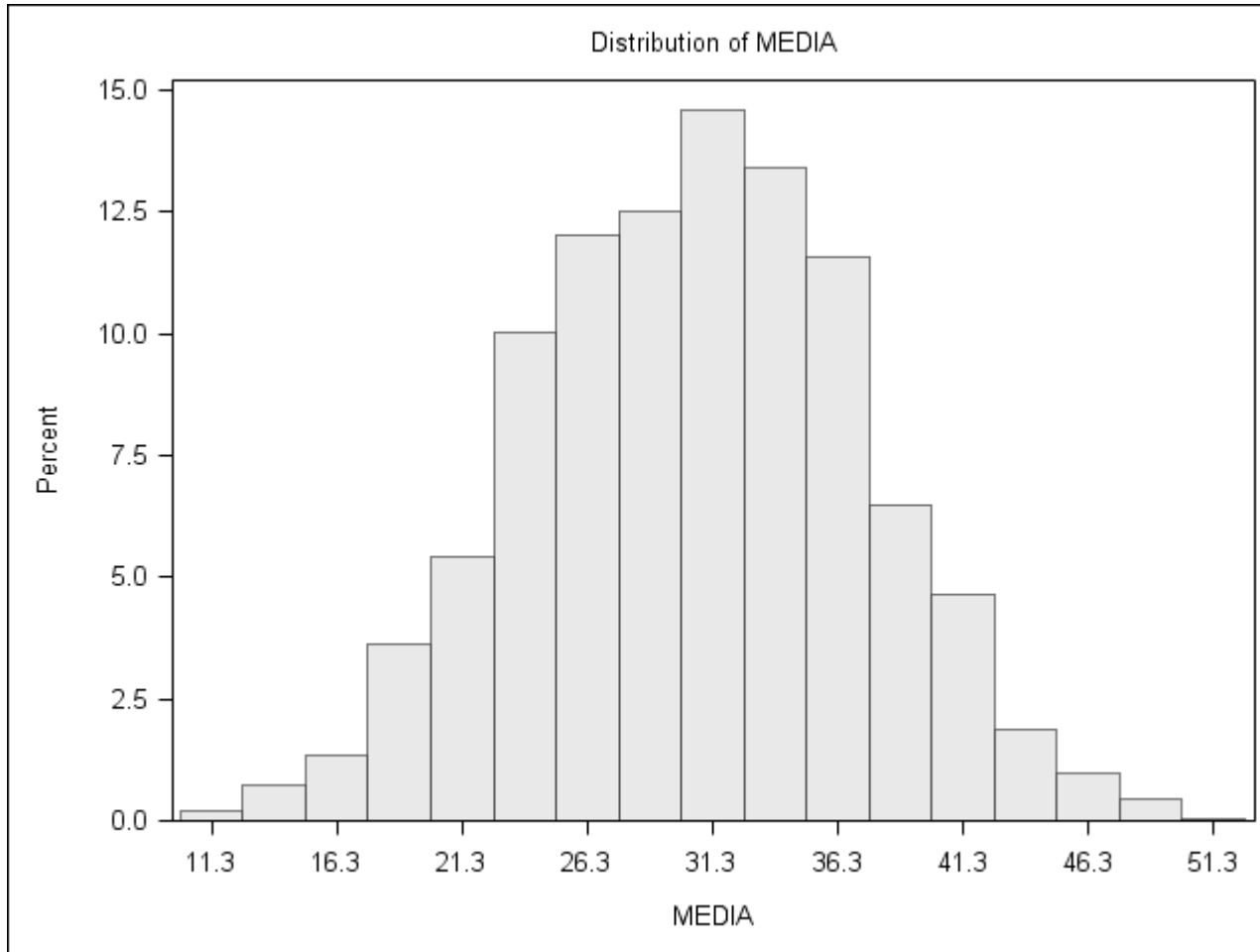
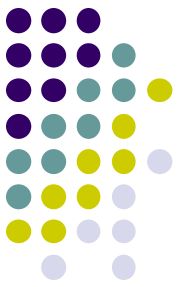




Tabela de números aleatórios

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| 1 | 25006 | 33852 | 24118 | 71965 | 18165 | 38827 | 46248 | 40197 | 64841 | 22440 |
| 2 | 62496 | 74877 | 12932 | 41986 | 08929 | 18405 | 61145 | 50954 | 07941 | 33036 |
| 3 | 28314 | 80808 | 53395 | 34515 | 24392 | 51518 | 24761 | 11013 | 80761 | 26216 |
| 4 | 72901 | 94543 | 38959 | 11994 | 66053 | 30153 | 87912 | 62162 | 26465 | 25192 |
| 5 | 24711 | 87263 | 96617 | 11449 | 88458 | 69602 | 88154 | 65563 | 04699 | 16928 |
| 6 | 06460 | 08533 | 39285 | 73260 | 17546 | 90405 | 43002 | 29025 | 61979 | 54174 |
| 7 | 55773 | 65127 | 53288 | 58669 | 89943 | 61428 | 29389 | 22638 | 77487 | 70358 |
| 8 | 13278 | 70775 | 93658 | 46679 | 90711 | 33136 | 37945 | 94574 | 81564 | 58825 |
| 9 | 41443 | 44690 | 14222 | 23279 | 90270 | 83237 | 24356 | 78350 | 25181 | 58285 |
| 10 | 20172 | 39205 | 43267 | 13158 | 67123 | 70489 | 24486 | 43570 | 98010 | 63326 |
| 11 | 19681 | 88247 | 20981 | 19433 | 21867 | 14729 | 04761 | 88318 | 74163 | 28004 |
| 12 | 94663 | 15205 | 94244 | 16606 | 51981 | 00573 | 00614 | 59473 | 63445 | 87235 |
| 13 | 95787 | 53252 | 18023 | 45891 | 32839 | 68650 | 47927 | 22267 | 93755 | 79023 |
| 14 | 45118 | 51214 | 36673 | 55099 | 41116 | 42439 | 35550 | 80604 | 31385 | 00018 |
| 15 | 75121 | 20134 | 75572 | 26736 | 56839 | 81898 | 20167 | 46000 | 38221 | 60167 |
| 16 | 55876 | 47702 | 51958 | 36837 | 92476 | 78797 | 11627 | 41142 | 91451 | 17124 |
| 17 | 36086 | 73477 | 27353 | 81682 | 57538 | 14587 | 23853 | 91612 | 01905 | 05582 |
| 18 | 04407 | 20284 | 80142 | 67756 | 31049 | 30440 | 12267 | 01868 | 53839 | 70174 |
| 19 | 37852 | 60206 | 16799 | 97634 | 24384 | 96802 | 88467 | 83693 | 64604 | 25317 |

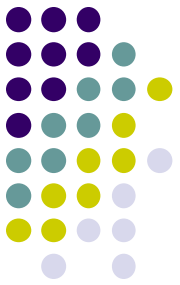
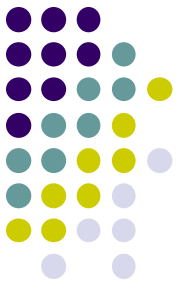


Tabela de números aleatórios na internet

**[HTTP://TEORICA.FIS.UCM.ES/FT8/
TABLERN2.PDF](http://teorica.fis.ucm.es/ft8/tablern2.pdf)**

Programa SAS para gerar números aleatórios:

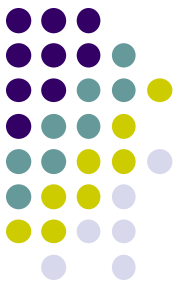


- Vamos supor que queremos 10 números de 1 a **100**.
- DATA A;
- DO NUMERO=1 TO **100**;
- OUTPUT;END;
- PROC PRINT DATA=A;
- RUN;
- PROC SURVEYSELECT DATA=A METHOD=**SRS** N=**10** OUT=B;
- PROC PRINT DATA=B;
- RUN;
- PROC SURVEYSELECT DATA=A METHOD=**SYS** N=**10** OUT=C;
- PROC PRINT DATA=C;
- RUN;

ALEATÓRIO

SISTEMÁTICO

Cálculo de estatísticas descritivas com SAS.



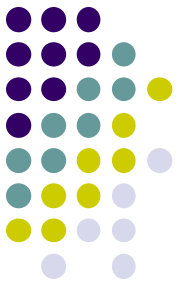
- O PROC MEANS é o procedimento usado para realizar diversos cálculos, incluindo a estatística descritiva. Outros Procs, como UNIVARIATE, SUMMARY também podem ser usados para calcular algumas estatísticas descritivas.
- Geralmente é usado com o PROC SORT, que ordena os dados por alguma variável nominal ou numérica.
- Usado também com os comandos SET ou MERGE.



Exemplo:

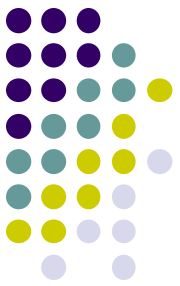
| Local | Ano | Precipitação(mm) | Temperatura (°C) |
|------------|------|------------------|------------------|
| Candeias | 2001 | 1256 | 26 |
| Candeias | 2002 | 1089 | 28 |
| Candeias | 2003 | 1152 | 27 |
| Catú | 2001 | 952 | 30 |
| Catú | 2002 | 894 | 29 |
| Catú | 2003 | 569 | 28 |
| Entre Rios | 2001 | 1350 | 31 |
| Entre Rios | 2002 | 1260 | 32 |
| Entre Rios | 2003 | 1420 | 30 |
| Amargoso | 2001 | 782 | 32 |
| Amargoso | 2002 | 695 | 33 |
| Amargoso | 2003 | 529 | 33 |

Programa SAS



```
DATA MM;  
INPUT LOCAL $ 1-10 ANO PRECIP TEMP;  
DATALINES;  
Candeias 2001 1256 26  
Candeias 2002 1089 28  
Candeias 2003 1152 27  
Catú 2001 952 30  
Catú 2002 894 29  
Catú 2003 569 28  
Entre Rios 2001 1350 31  
Entre Rios 2002 1260 32  
Entre Rios 2003 1420 30  
Amargoso 2001 782 32  
Amargoso 2002 695 33  
Amargoso 2003 529 33  
;;;
```

Programa SAS 2



```
PROC SORT DATA=MM;
  BY LOCAL;
PROC MEANS DATA=MM NOPRINT;
  BY LOCAL;
VAR PRECIP TEMP;
OUTPUT OUT=MM1 MIN=MIN_MM MIN_TEM MAX=MAX_MM MAX_TEM MEAN=MED_MM
MED_TEM CV=CV_MM CV_TEM;
RUN;
PROC SORT DATA=MM1;
  BY DESCENDING MED_MM;
PROC PRINT DATA=MM1 NOOBS LABEL SPLIT='*';
  VAR LOCAL MIN_MM MAX_MM MED_MM CV_MM;
  LABEL LOCAL='Local*de*Coleta'
  MIN_MM='Precipitação*Mínima*(mm) '
  MAX_MM='Precipitação*Máxima*(mm) '
  MED_MM='Precipitação*Média*(mm) '
  CV_MM='Coeficiente*de*Variação*(%)';
  FORMAT MIN_MM MAX_MM COMMAX6.0 MED_MM CV_MM COMMAX8.1;
RUN;
PROC SORT DATA=MM1;
  BY DESCENDING MED_TEM;
PROC PRINT DATA=MM1 NOOBS LABEL SPLIT='*';
  VAR LOCAL MIN_TEM MAX_TEM MED_TEM CV_TEM;
  LABEL LOCAL='Local*de*Coleta'
  MIN_TEM='Temperatura*Mínima*(°C) '
  MAX_TEM='Temperatura*Máxima*(°C) '
  MED_TEM='Temperatura*Média*(°C) '
  CV_TEM='Coeficiente*de*Variação*(%)';
  FORMAT MIN_TEM MAX_TEM COMMAX6.0 MED_TEM CV_TEM COMMAX6.1;
RUN;
```

Resultado:



| Local de Coleta | Precipitação Mínima (mm) | Precipitação Máxima (mm) | Precipitação Média (mm) | Coefficiente de Variação (%) |
|-----------------|--------------------------|--------------------------|-------------------------|------------------------------|
| Entre Rios | 1.260 | 1.420 | 1.343,3 | 6,0 |
| Candeias | 1.089 | 1.256 | 1.165,7 | 7,2 |
| Catú | 569 | 952 | 805,0 | 25,6 |
| Amargoso | 529 | 782 | 668,7 | 19,2 |

| Local de Coleta | Temperatura Mínima (°C) | Temperatura Máxima (°C) | Temperatura Média (°C) | Coefficiente de Variação (%) |
|-----------------|-------------------------|-------------------------|------------------------|------------------------------|
| Amargoso | 32 | 33 | 32,7 | 1,8 |
| Entre Rios | 30 | 32 | 31,0 | 3,2 |
| Catú | 28 | 30 | 29,0 | 3,4 |
| Candeias | 26 | 28 | 27,0 | 3,7 |

OBRIGADO !!!



*ATÉ A
PRÓXIMA !!!*