JORNALISMO DE DADOS E VISUALIZAÇÃO

Noções de Estatística para Jornalistas

Marcelo Leme de Arruda
www.chancedegol.com.br

Introdução

Conceitos matemáticos

1 – Somatório (Σ)

Soma geral de termos

Notação:

$$\sum_{i=1}^{n} a_i$$

soma dos valores a_i para i variando de 1 até n.

<u>Introdução</u>

Conceitos matemáticos

1 – Somatório (Σ)

Exemplos:

$$\sum_{i=1}^{8} X_i = X_1 + X_2 + X_3 + X_4 + X_5 + X_6 + X_7 + X_8$$

$$\sum_{j=1}^{5} 3j = 3 + 6 + 9 + 12 + 15$$

$$\sum_{n=2}^{4} (n^2 + n) = (4+2) + (9+3) + (16+4)$$

<u>Introdução</u>

Conceitos matemáticos

2 – Produtório (Π)

Produto geral de termos

Notação:

$$\prod_{i=1}^{n} a_i$$

produto dos valores a_i para *i* variando de 1 até *n*.

<u>Introdução</u>

Conceitos matemáticos

2 – Produtório (Π)

Exemplos:

$$\prod_{i=1}^{4} X_i = X_1 \cdot X_2 \cdot X_3 \cdot X_4$$

$$\prod_{k=1}^{3} (k+1) = 2 \cdot 3 \cdot 4$$

$$\prod_{j=3}^{6} \frac{j^3}{j-1} = \frac{3^3}{2} \cdot \frac{4^3}{3} \cdot \frac{5^3}{4} \cdot \frac{6^3}{5}$$

Introdução

Conceitos matemáticos

3 – Fatorial (!)

Produto de todos os números de 1 até n

Definição:

$$n! = \prod_{i=1}^{n} i = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot (n-1) \cdot n$$

Introdução

Conceitos matemáticos

3 – Fatorial (!)

Exemplos:

$$5! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5$$

$$9! = 1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8 \cdot 9$$

por definição: $0! = 1$

<u>Introdução</u>

Conceitos matemáticos

4 – Permutação (P)

De quantas maneiras se pode ordenar (permutar) um conjunto de *k* elementos?

Resposta:

* Temos *k* escolhas possíveis para o primeiro elemento (i.e. *k* elementos possíveis).

* Para cada uma dessas *k* alternativas, temos *k-1* escolhas possíveis para o segundo elemento (i.e. *k(k-1)* pares possíveis).

<u>Introdução</u>

Conceitos matemáticos

4 – Permutação (P)

De quantas maneiras se pode ordenar (permutar) um conjunto de *k* elementos?

Resposta:

* Para cada uma dessas $k(k-1)$ alternativas, temos $k-2$ escolhas possíveis para o terceiro elemento (i.e. $k(k-1)(k-2)$ trios possíveis)

* e assim por diante, até o último elemento, para o qual só há uma escolha possível.

<u>Introdução</u>

Conceitos matemáticos

4 – Permutação (P)

De quantas maneiras se pode ordenar (permutar) um conjunto de *k* elementos?

Resposta:

logo, $P_k = k \cdot (k-1) \cdot (k-2) \cdot \ldots \cdot 1 = k!$

Permutação - Exemplo:

elementos: letras A, B, C e D

* 4 alternativas para a primeira letra:

A	B
C	D

total de alternativas:

4

Permutação - Exemplo:

elementos: letras A, B, C e D

* Para cada uma dessas possibilidades, temos 3 alternativas para a segunda letra:

AB	BA	
AC	BC	total de alternativas:
AD	BD	4 x 3
CA	DA	
CB	DB	
CD	DC	

Permutação - Exemplo:

elementos: letras A, B, C e D

* Para cada uma dessas possibilidades, temos 2 alternativas para a terceira letra:

ABC	ABD	BAC	BAD
ACB	ACD	BCA	BCD
ADB	ADC	BDA	BDC
CAB	CAD	DAB	DAC
CBA	CBD	DBA	DBC
CDA	CDB	DCA	DCB

total de alternativas:

$4 \times 3 \times 2$

Permutação - Exemplo:

elementos: letras A, B, C e D

* Para cada uma dessas possibilidades, temos 1 alternativa para a quarta letra:

ABCD	ABDC	BACD	BADC
ACBD	ACDB	BCAD	BCDA
ADBC	ADCB	BDAC	BDCA
CABD	CADB	DABC	DACB
CBAD	CBDA	DBAC	DBCA
CDAB	CDBA	DCAB	DCBA

total de alternativas:

$4 \times 3 \times 2 \times 1 = 4!$

<u>Introdução</u>

Conceitos matemáticos

5 – Combinação

De um conjunto de *n* elementos, de quantas maneiras pode-se escolher *k* deles?

Notação:

$$C_{n,k} = \binom{n}{k}$$

combinação de *n* elementos, tomados (escolhidos) *k*

Combinação - Exemplo:

Mega Sena: $n = 50$ e $k = 6$

Então temos:

* 50 possibilidades para o primeiro número sorteado

* 49 possibilidades para o segundo número sorteado

* 48 possibilidades para o terceiro número sorteado

* 47 possibilidades para o quarto número sorteado

* 46 possibilidades para o quinto número sorteado

* 45 possibilidades para o sexto número sorteado

Combinação - Exemplo:

Mega Sena: $n = 50$ e $k = 6$

Portanto, teríamos, a princípio:

50 x 49 x 48 x 47 x 46 x 45 possibilidades

MAS:

(25, 14, 19, 38, 07, 50) foi contabilizada

(14, 38, 50, 07, 25, 19) também foi contabilizada

(38, 25, 14, 50, 19, 07) também foi contabilizada

e assim por diante - e todas essas possibilidades são, na realidade, a mesma possibilidade!

Combinação - Exemplo:

Mega Sena: *n = 50* e *k = 6*

Precisamente, cada possibilidade foi contabilizada

P_6 = 6! vezes (o número de permutações possíveis

de um conjunto de 6 elementos).

Logo, o número correto de possibilidades diferentes de selecionarmos 6 números dentre os 50 da Mega Sena é:

$$C_{50,6} = \frac{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45}{6!}$$

Combinação - Exemplo:

Mega Sena: *n = 50* e *k = 6*

Essa fórmula pode ficar mais fácil de ser manipulada se multiplicarmos numerador e denominador da seguinte forma:

$$C_{50,6} = \frac{50 \cdot 49 \cdot 48 \cdot 47 \cdot 46 \cdot 45}{6!} \cdot \frac{44 \cdot 43 \cdot \ldots \cdot 2 \cdot 1}{44 \cdot 43 \cdot \ldots \cdot 2 \cdot 1}$$

e chegarmos a:

$$C_{50,6} = \frac{50!}{6! \cdot 44!}$$

<u>Introdução</u>

Conceitos matemáticos

5 – Combinação

De um conjunto de *n* elementos, de quantas maneiras pode-se escolher *k* deles?

Resposta: generalizando o exemplo anterior, podemos chegar à fórmula

$$C_{n,k} = \binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Introdução

No Excel:

=SOMA(...) – soma valores de uma área da planilha

=MULT(...) – multiplica valores de uma área

=FATORIAL(k) – calcula k!

=COMBIN(n;k) – calcula $C_{n,k}$

ATENÇÃO: o Excel possui uma função PERMUT, mas ela calcula o número de arranjos (e não de permutações).

<u>I – Estatística Descritiva</u>

Medidas que <u>descrevem</u> e resumem características relevantes de um conjunto de dados. Se dividem em:

* Medidas de Tendência Central

Medidas relativas a localização, posição, ordem de grandeza dos dados.

* Medidas de Dispersão

Medidas relativas a como os dados se distribuem (se <u>dispersam</u>) ou se concentram.

I – Estatística Descritiva

* Medidas de Tendência Central

1 – Média (aritmética)

Notação: $\overline{X} = \dfrac{\sum_{i=1}^{n} X_i}{n}$

onde $X_1, X_2, X_3, \cdots, X_{n-1}, X_n$ representam os dados que estão sendo analisados e *n* é a quantidade de dados em questão.

I – Estatística Descritiva

* Medidas de Tendência Central

2 – Mediana

Valor que divide os dados ao meio, ou seja, deixando exatamente 50% dos dados acima e 50% abaixo dele.

Isso significa que:

a) Quando *n* é impar, a mediana é o valor central dos dados ordenados;

b) Quando *n* é par, a mediana é a média dos dois valores centrais dos dados ordenados.

I – Estatística Descritiva

* Medidas de Tendência Central

3 – Moda

Valor que aparece com maior freqüência no conjunto de dados.

Observações:

a) Quando existe mais de um valor mais freqüente, diz-se que a distribuição dos dados é bimodal, trimodal, multimodal etc.

b) Quando os dados são trabalhados de forma agrupada, fala-se em faixa modal, grupo modal etc.

<u>I – Estatística Descritiva</u>

* Medidas de Tendência Central

Exemplo 1:

Dados: 2 1 2 6 3

então: $n = 5$

$X_1 = 2$ $\quad X_2 = 1$ $\quad X_3 = 2$

$X_4 = 6$ $\quad X_5 = 3$

dados ordenados: 1 2 2 3 6

I – Estatística Descritiva

* Medidas de Tendência Central

Exemplo 1:

Média: $\overline{X} = \dfrac{2+1+2+6+3}{5} = 2{,}8$

Mediana = 2 (1 2 2 3 6)

Moda = 2 (1 2 2 3 6)

I – Estatística Descritiva

* Medidas de Tendência Central

Exemplo 2:

Dados: 3 1 4 6 4 3

então: $n = 6$

$X_1 = 3$ $X_2 = 1$ $X_3 = 4$

$X_4 = 6$ $X_5 = 4$ $X_6 = 3$

dados ordenados: 1 3 3 4 4 6

I – Estatística Descritiva

* Medidas de Tendência Central

Exemplo 2:

Média: $\bar{X} = \dfrac{3+1+4+6+4+3}{6} = 3{,}5$

Mediana = $\dfrac{3+4}{2} = 3{,}5$ (1 3 3 4 4 6)

Moda = distribuição bimodal (1 3 3 4 4 6)

I – Estatística Descritiva

* Medidas de Tendência Central

No Excel:

=MÉDIA(...) – média dos valores de uma área

=MED(...) – mediana dos valores de uma área

=MODO(...) – moda dos valores de uma área

ATENÇÃO: para conjuntos de dados multimodais, a função MODO retorna somente uma das modas.

I – Estatística Descritiva

* Observação: tipos de dados

a) Qualitativos (exemplo: time de preferência): só permite moda.

b) Ordinais (exemplo: avaliação de satisfação): moda e mediana.

DADOS NÃO NUMÉRICOS

c) Quantitativos Discretos (exemplo: idade): moda (número), mediana e média.

d) Quantitativos Contínuos (exemplo: IMC): moda (faixa), mediana e média.

DADOS NUMÉRICOS

I – Estatística Descritiva

* Medidas de Dispersão

1 – Variância

Notação: $S^2 = \dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}$

interpretação: média das distâncias (quadráticas) de cada valor à média.

obs: às vezes usa-se a notação σ^2 em vez de S^2.

I – Estatística Descritiva

* Medidas de Dispersão

2 – Desvio Padrão

Notação: $S = \sqrt{S^2} = \sqrt{\dfrac{\sum_{i=1}^{n}(X_i - \overline{X})^2}{n-1}}$

obs: O desvio padrão segue a mesma unidade de medida que a média (isso não acontece com a variância).

I – Estatística Descritiva

* Medidas de Dispersão

3 – Amplitude

Notação: $\Delta = Max(X_i) - Min(X_i)$

onde:

$Max(X_i)$ = maior valor dos dados

$Min(X_i)$ = menor valor dos dados

<u>I – Estatística Descritiva</u>

* Medidas de Dispersão

4 – Intervalo Interquartil

Notação: $IIQ = Q_3 - Q_1$

onde:

Q_3 (<u>3º quartil</u>) = Valor que divide os dados deixando exatamente 25% dos dados acima e <u>75%</u> abaixo dele.

Q_1 (<u>1º quartil</u>) = Valor que divide os dados deixando exatamente 75% dos dados acima e <u>25%</u> abaixo dele.

<u>I – Estatística Descritiva</u>

* Medidas de Dispersão

4 – Intervalo Interquartil

Notação: $IIQ = Q_3 - Q_1$

observações:

a) Valem para os quartis observações análogas às feitas para a mediana.

b) O IIQ e a mediana são mais <u>robustos</u> que a amplitude, o desvio padrão e a média

<u>I – Estatística Descritiva</u>

* Medidas de Dispersão

Exemplo 3:

Dados: 2 3 5 7 7 5 2 9

então: $n = 8$

$X_1 = 2$	$X_2 = 3$	$X_3 = 5$	$X_4 = 7$
$X_5 = 7$	$X_6 = 5$	$X_7 = 2$	$X_8 = 9$

dados ordenados: 2 2 3 5 5 7 7 9

I – Estatística Descritiva

* Medidas de Dispersão

Exemplo 3:

Média: $\overline{X} = \dfrac{2+3+5+7+7+5+2+9}{8} = 5$

Variância: $S^2 = \dfrac{(2-5)^2+(3-5)^2+(5-5)^2+(7-5)^2+}{8}$

$\dfrac{+(7-5)^2+(5-5)^2+(2-5)^2+(9-5)^2}{8} = 5{,}75$

Desvio Padrão: $S = \sqrt{5{,}75} = 2{,}398$

I – Estatística Descritiva

* Medidas de Dispersão

Exemplo 3:

Amplitude: $\Delta = 9 - 2 = 7$

$Q_1 = \dfrac{2+3}{2} = 2,5$ (2 2 3 5 5 7 7 9)

$Q_3 = \dfrac{7+7}{2} = 7$ (2 2 3 5 5 7 7 9)

Intervalo Interquartil: $IIQ = 7 - 2,5 = 4,5$

I – Estatística Descritiva

* Medidas de Dispersão

No Excel:

=VAR(...) – variância dos valores de uma área

=DESVPAD(...) – desvio padrão desses valores

ATENÇÃO: Existem também as funções VARP e DESVPADP, que fazem esses cálculos com o denominador *n*.

<u>I – Estatística Descritiva</u>

* Medidas de Dispersão

No Excel:

=MÁXIMO(...) – máximo dos valores de uma área

=MÍNIMO(...) – mínimo dos valores de uma área

=QUARTIL(...;3) – 3º quartil desses valores

=QUARTIL(...;1) – 1º quartil desses valores

I – Estatística Descritiva

* Outras Medidas

1 – Coeficiente de Variação

Notação: $CV = \dfrac{S}{\overline{X}}$

interpretação: "grandeza" do desvio padrão em relação à média.

<u>I – Estatística Descritiva</u>

* Outras Medidas

2 - Assimetria

Notação: $$Ass = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^3}{n}$$

interpretação:

Ass = 0 $\Rightarrow$ distribuição simétrica

Ass < 0 $\Rightarrow$ distribuição assimétrica à esquerda

Ass > 0 $\Rightarrow$ distribuição assimétrica à direita

I – Estatística Descritiva

* Outras Medidas

2 - Assimetria

Notação: $Ass = \dfrac{\sum_{i=1}^{n} (X_i - \overline{X})^3}{n}$

de um modo geral (para valores "grandes" de *n*):

Ass = 0 $\Rightarrow$ média = mediana = moda

Ass < 0 $\Rightarrow$ média < mediana < moda

Ass > 0 $\Rightarrow$ média > mediana > moda

I – Estatística Descritiva

* Outras medidas

Exemplo 4:

X_1 = 11
X_2 = 12
X_3 = 13
X_4 = 15
X_5 = 15
X_6 = 15
X_7 = 15
X_8 = 22
X_9 = 26

$Ass(X) = 996$

(assimetria à direita)

$\overline{X} = 16$ (> mediana)

I – Estatística Descritiva

* Outras medidas

Exemplo 4:

$Y_1 = 12$
$Y_2 = 13$
$Y_3 = 14$
$Y_4 = 15$
$Y_5 = 15$
$Y_6 = 15$
$Y_7 = 16$
$Y_8 = 17$
$Y_9 = 18$

$Ass(Y) = 0$

(simetria)

$\bar{Y} = 15$ (= mediana)

I – Estatística Descritiva

* Outras medidas

Exemplo 4:

Z_1 = 3
Z_2 = 4
Z_3 = 15
Z_4 = 15
Z_5 = 15
Z_6 = 15
Z_7 = 16
Z_8 = 17
Z_9 = 17

$Ass(Z) = -1542$

(assimetria à esquerda)

$\overline{Z} = 13$ (< mediana)

I – Estatística Descritiva

* Outras Medidas

No Excel:

=DISTORÇÃO(...) – assimetria dos valores de uma área

ATENÇÃO: A função DISTORÇÃO utiliza uma fórmula um pouco diferente da mostrada nos slides anteriores, mas a interpretação do resultado (em função do sinal) permanece a mesma.

I – Estatística Descritiva

* Medidas de relação entre variáveis

1 – Covariância

Notação: $$Cov_{X,Y} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{n}$$

interpretação:

Cov < 0 $\Rightarrow$ quanto maior é o valor de X, menor é o de Y (i.e. uma variável cresce à medida que outra decresce).

exemplo: gols sofridos x pontos ganhos

I – Estatística Descritiva

* Medidas de relação entre variáveis

1 – Covariância

Notação: $$Cov_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n}$$

interpretação:

Cov > 0 $\Rightarrow$ quanto maior é o valor de X, maior é o de Y (i.e. uma variável cresce à medida que outra decresce).

exemplo: altura x nº do sapato

<u>I – Estatística Descritiva</u>

* Medidas de relação entre variáveis

1 – Covariância

Notação: $$Cov_{X,Y} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{n}$$

interpretação:

Cov = 0 $\Rightarrow$ ausência de relação

exemplo: nº de letras do sobrenome x último dígito do RG

I – Estatística Descritiva

* Medidas de relação entre variáveis

2 – Correlação

Notação: $\rho_{X,Y} = \dfrac{Cov_{X,Y}}{S_X S_Y}$

obs: a) interpretação igual à da covariância, com o diferencial de que $\rho_{X,Y}$ está sempre entre -1 e 1.

b) Aqui usa-se os desvios padrão S_X e S_Y calculados com denominador n (e não $n-1$).

I – Estatística Descritiva

* Medidas de relação entre variáveis

Exemplo 4:

Dados:

$X_1 = 6$ $X_2 = 7$ $X_3 = 8$

$Y_1 = 10$ $Y_2 = 8$ $Y_3 = 6$

$n = 3$

$$\bar{X} = \frac{6+7+8}{3} = 7 \qquad \bar{Y} = \frac{10+8+6}{3} = 8$$

I – Estatística Descritiva

* Medidas de relação entre variáveis

Exemplo 4:

Dados:

$X_1 = 6$ $X_2 = 7$ $X_3 = 8$

$Y_1 = 10$ $Y_2 = 8$ $Y_3 = 6$

$$S_X = \sqrt{\frac{(6-7)^2 + (7-7)^2 + (8-7)^2}{3}} = \sqrt{\frac{2}{3}}$$

$$S_Y = \sqrt{\frac{(10-8)^2 + (8-8)^2 + (6-8)^2}{3}} = \sqrt{\frac{8}{3}}$$

I – Estatística Descritiva

* Medidas de relação entre variáveis

Exemplo 4:

Dados:

$X_1 = 6$	$X_2 = 7$	$X_3 = 8$
$Y_1 = 10$	$Y_2 = 8$	$Y_3 = 6$

$$Cov_{X,Y} = \frac{(6-7)(10-8)+(7-7)(8-8)+(8-7)(6-8)}{3} = -\frac{4}{3}$$

$$\rho_{X,Y} = \frac{-4/3}{\sqrt{2/3}\cdot\sqrt{8/3}} = -1$$

I – Estatística Descritiva

* Medidas de relação entre variáveis

No Excel:

=COVAR(...;...) – covariância entre os valores armazenados em duas áreas

=CORREL(...;...) – correlação entre os valores armazenados em duas áreas

II – Probabilidades

*** Conceito geral:**

$$probabilidade = \frac{\#\text{casos favoráveis}}{\#\text{casos totais}}$$

Exemplo:

Arremesso de um dado comum

$$P(\text{sair um número par}) = \frac{\#\{2,4,6\}}{\#\{1,2,3,4,5,6\}} = \frac{3}{6} = \frac{1}{2}$$

<u>II – Probabilidades</u>

* Conceito geral:

$$probabilidade = \frac{\#\text{casos favoráveis}}{\#\text{casos totais}}$$

MAS, para a maioria dos "problemas reais", essa abordagem é ineficiente para calcular as probabilidades desejadas.

Isso motiva o uso de **Distribuições** (ou seja, de funções que, para cada valor possível, atribuem a probabilidade de obtermos esse valor).

<u>II – Probabilidades</u>

* Ensaio de Bernoulli:

Um Ensaio de Bernoulli é a realização de um evento cujo resultado pode ou não corresponder àquele em que temos interesse.

Quando corresponde, dizemos que houve um "sucesso"; quando não corresponde, dizemos que houve um "fracasso".

Exemplos: ensaio - arremesso de moeda
sucesso - sair cara
fracasso - sair coroa

II – Probabilidades

* Ensaio de Bernoulli:

Um Ensaio de Bernoulli é a realização de um evento cujo resultado pode ou não corresponder àquele em que temos interesse.

Quando corresponde, dizemos que houve um "sucesso"; quando não corresponde, dizemos que houve um "fracasso".

Exemplos: ensaio - sorteio de loteria
sucesso - sair um dos meus números
fracasso - sair um número diferente

II – Probabilidades

* Ensaio de Bernoulli:

Um Ensaio de Bernoulli é a realização de um evento cujo resultado pode ou não corresponder àquele em que temos interesse.

Quando corresponde, dizemos que houve um "sucesso"; quando não corresponde, dizemos que houve um "fracasso".

Exemplos: ensaio - sorteio de bolas de uma urna
sucesso - sair uma bola branca
fracasso - sair uma bola preta

II – Probabilidades

* Ensaio de Bernoulli:

Um Ensaio de Bernoulli é a realização de um evento cujo resultado pode ou não corresponder àquele em que temos interesse.

Quando corresponde, dizemos que houve um "sucesso"; quando não corresponde, dizemos que houve um "fracasso".

Notação: P(sucesso) = p
P(fracasso) = $1 - p$

II – Probabilidades

* Bolas e urnas; com ou sem reposição:

Essencialmente, qualquer Ensaio de Bernoulli pode ser representado por um sorteio de bolas de uma urna, podendo esse sorteio ser de dois tipos:

Ensaios com reposição: após o sorteio, a urna SEMPRE é reconstituída à sua composição original

Ensaios sem reposição: após o sorteio, a urna NUNCA é reconstituída à sua composição original

<u>II – Probabilidades</u>

* Bolas e urnas; com ou sem reposição:

<u>Exemplos</u>

1 – Arremessos de moeda:

Cada arremesso equivale um sorteio de uma urna com duas bolas, sendo uma branca (representando a cara) e uma preta (representando a coroa).

Após cada arremesso, a "urna" volta a ter uma bola branca (cara) e uma bola preta (coroa).

Logo, tratam-se de ensaios <u>com reposição</u>.

II – Probabilidades

* Bolas e urnas; com ou sem reposição:

Exemplos

2 – Sorteio da Mega Sena:

O sorteio da primeira dezena equivale a sortear de uma urna com 50 bolas, sendo x bolas brancas (os números da minha aposta) e $50 - x$ bolas pretas (todos os outros números possíveis).

Após essa dezena ser sorteada, a "urna" fica com 49 bolas e o total de brancas ou pretas também diminui em função da cor da bola sorteada.

Logo, tratam-se de ensaios sem reposição.

II – Probabilidades

* Distribuições baseadas em Bernoulli

Distribuição Binomial

É a distribuição do número de sucessos em *n* Ensaios de Bernoulli com reposição:

$$P(X = x) = \underbrace{\binom{n}{x}}_{\text{De quantas maneiras podemos ter } x \text{ sucessos em } n \text{ ensaios.}} \underbrace{p^x}_{\text{probabilidade de } x \text{ sucessos}} \underbrace{(1-p)^{n-x}}_{\text{probabilidade de } (n-x) \text{ fracassos}}$$

II – Probabilidades

* Distribuições baseadas em Bernoulli

Distribuição Binomial

Exemplo: probabilidade de observarmos 4 caras em 6 arremessos de uma moeda com P(cara) = *p* = 0,40

$$P(X = x) = \binom{6}{4}(0,4)^4(0,6)^2 = 0,138$$

No Excel:

=DISTRBINOM(x;n;p;0) – probabilidade de X = x

=DISTRBINOM(x;n;p;1) – probabilidade de X $\leq$ x

II – Probabilidades

* Distribuições baseadas em Bernoulli

Distribuição Hipergeométrica

É a distribuição do número de sucessos em *n* Ensaios de Bernoulli sem reposição, considerando que a urna tem inicialmente *N* bolas ao todo, sendo *k* brancas.

De quantas maneiras podemos sortear *x* bolas brancas de um total de *k*.

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}$$

De quantas maneiras podemos sortear *n-x* bolas pretas de um total de *N-k*.

De quantas maneiras podemos sortear *n* bolas de um total de *N*.

II – Probabilidades

* Distribuições baseadas em Bernoulli

Distribuição Hipergeométrica

Exemplo: probabilidade de, apostando oito dezenas na mega sena, acertarmos cinco números sorteados.

$$P(X = x) = \frac{\binom{8}{5}\binom{42}{1}}{\binom{50}{6}} = 0{,}000148$$

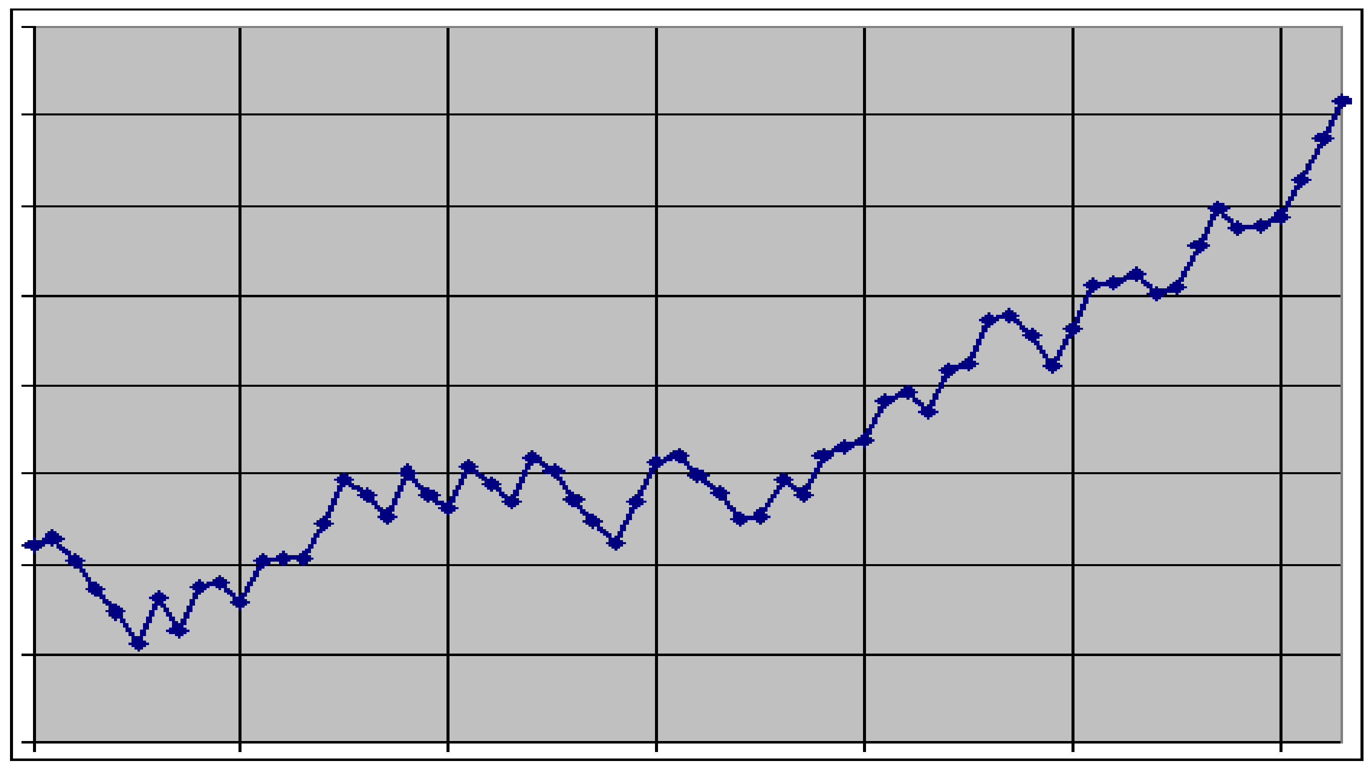
No Excel:

=DIST.HIPERGEOM(x;k;n;N)

III – Gráficos

* Gráfico de Dispersão

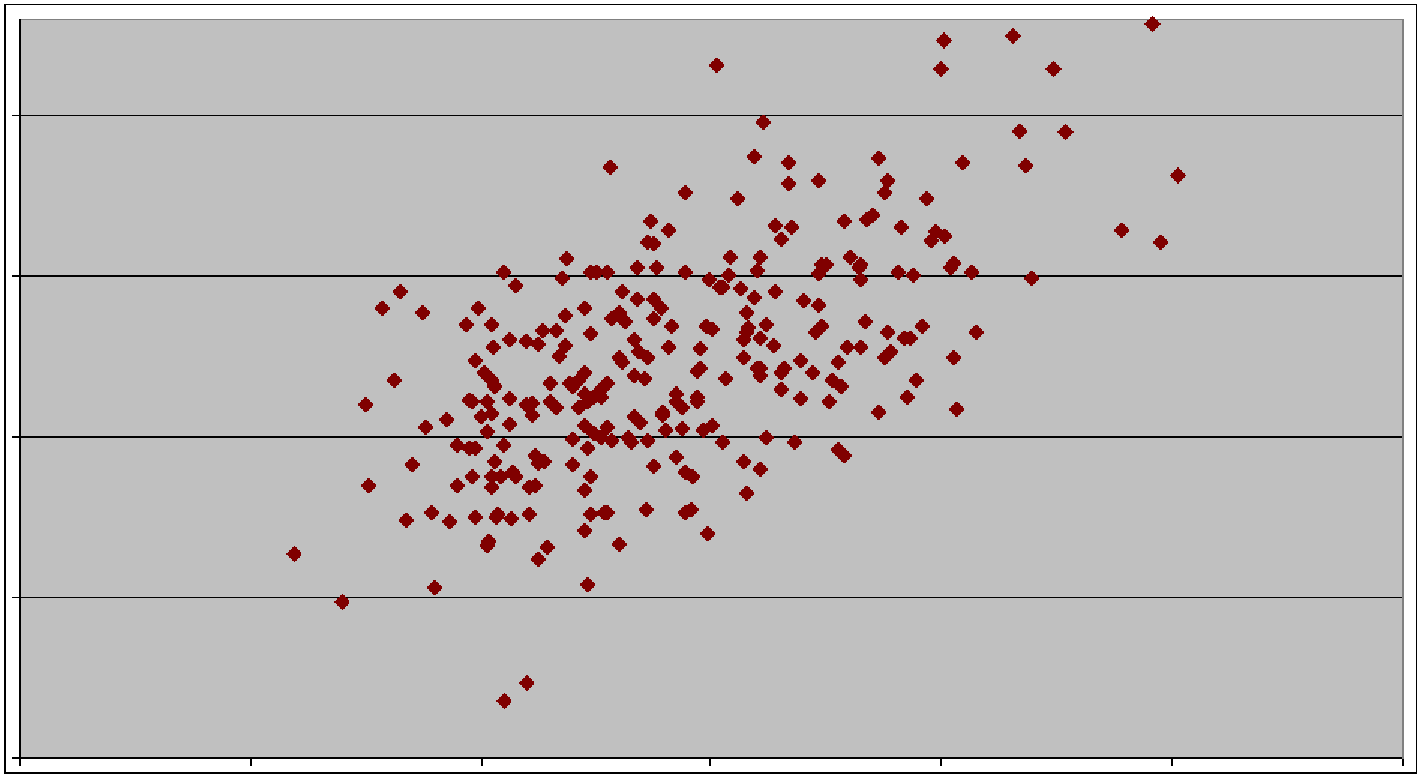
Exemplo: evolução de um índice ao longo do tempo



III – Gráficos

* Gráfico de Dispersão

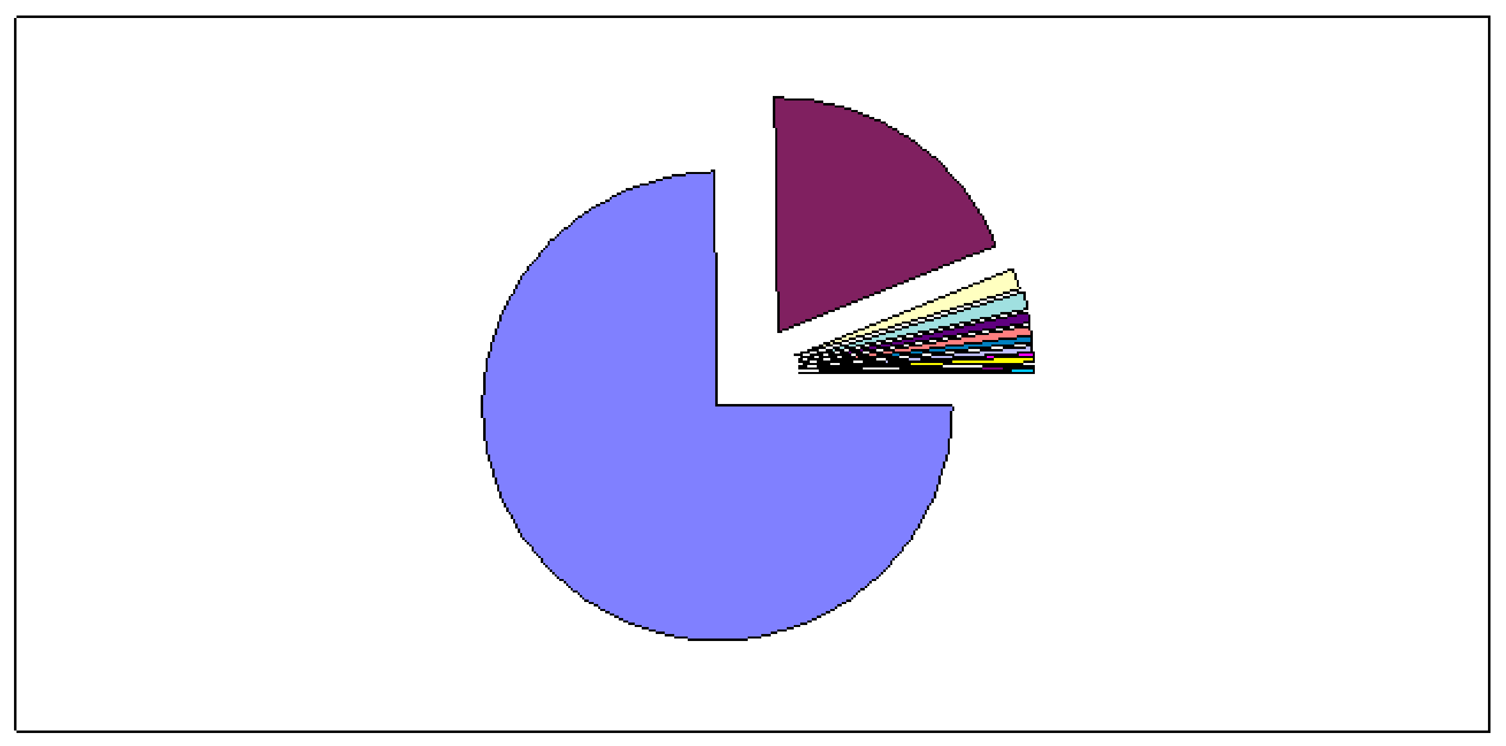
Exemplo: notas de alunos em duas provas



III – Gráficos

* Gráfico de Setores ("Pizza")

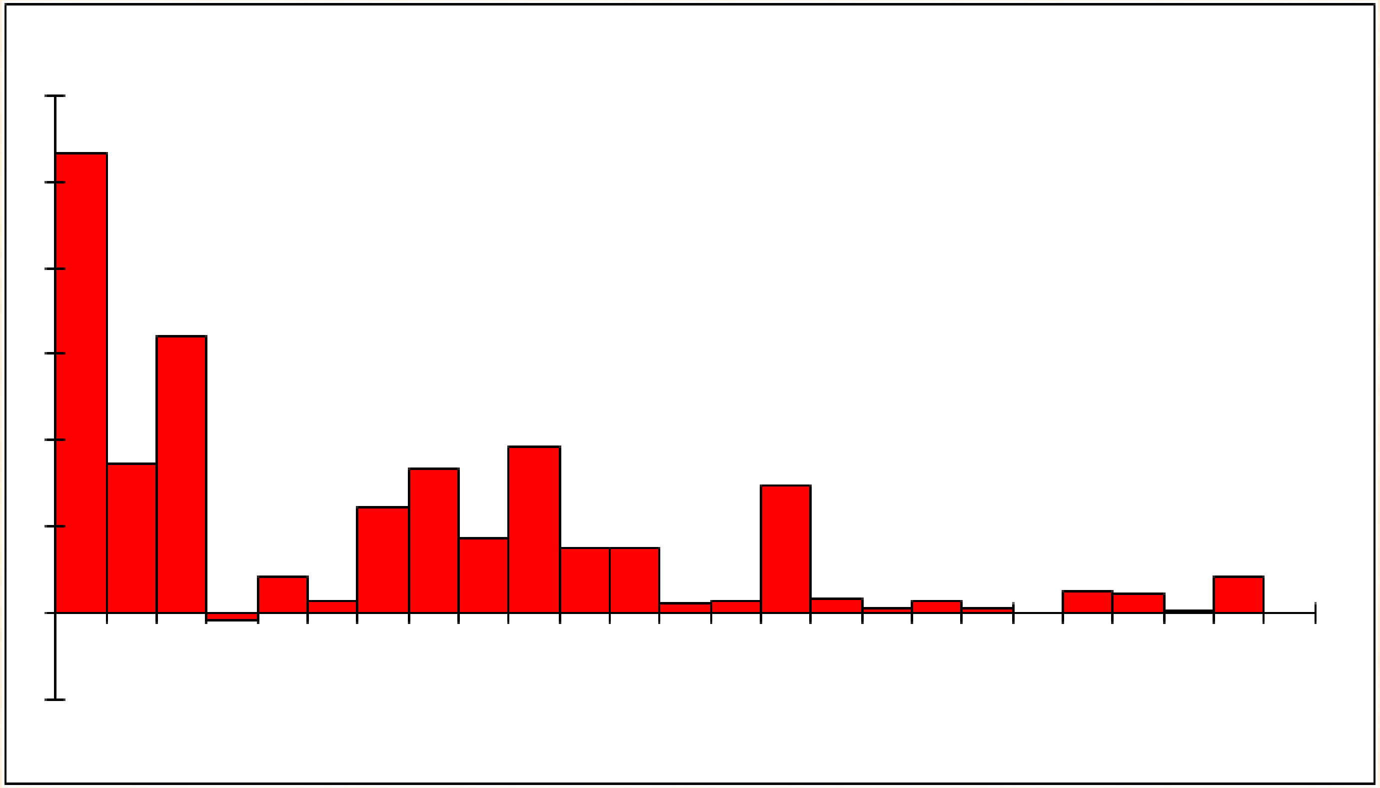
Exemplo: consumo de marcas de um produto



III – Gráficos

* Gráfico de Barras

Exemplo: aprovação ao governo x partido de preferência



III – Gráficos

* Gráficos Pictóricos

Exemplo:

Nascimentos em 1993:

5

Nascimentos em 2003:

20

Nascimentos em 2003:

68

III – Gráficos

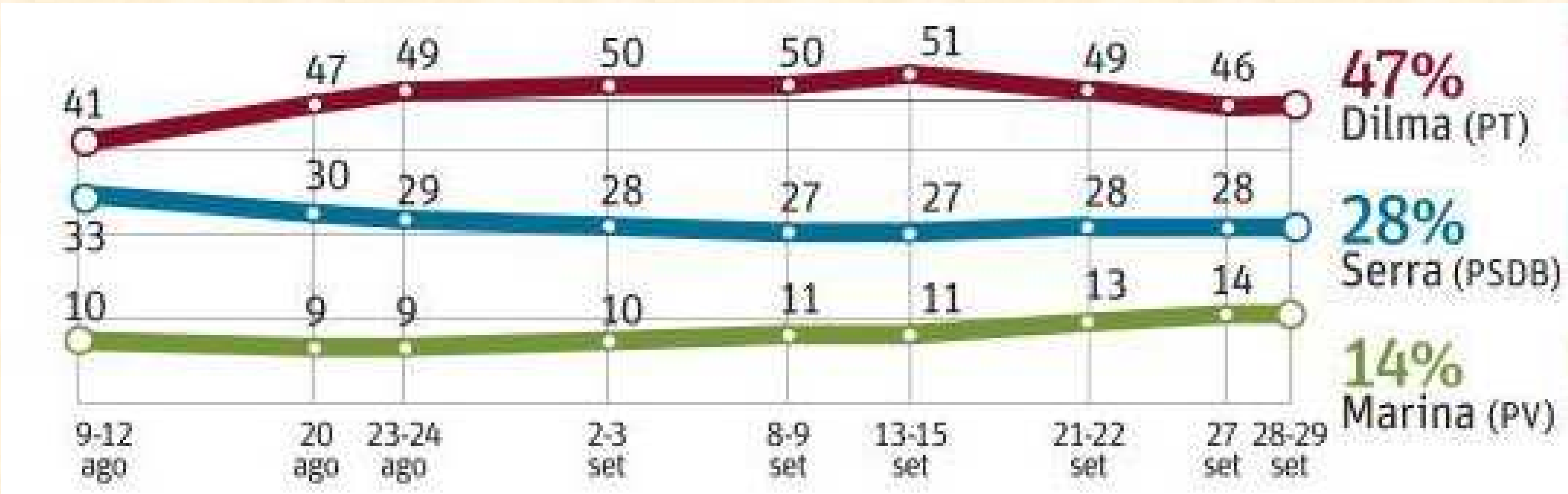
* Observações e cuidados

a) Gráficos de Dispersão: ESCALA

III – Gráficos

* Observações e cuidados

a) Gráficos de Dispersão: ESCALA

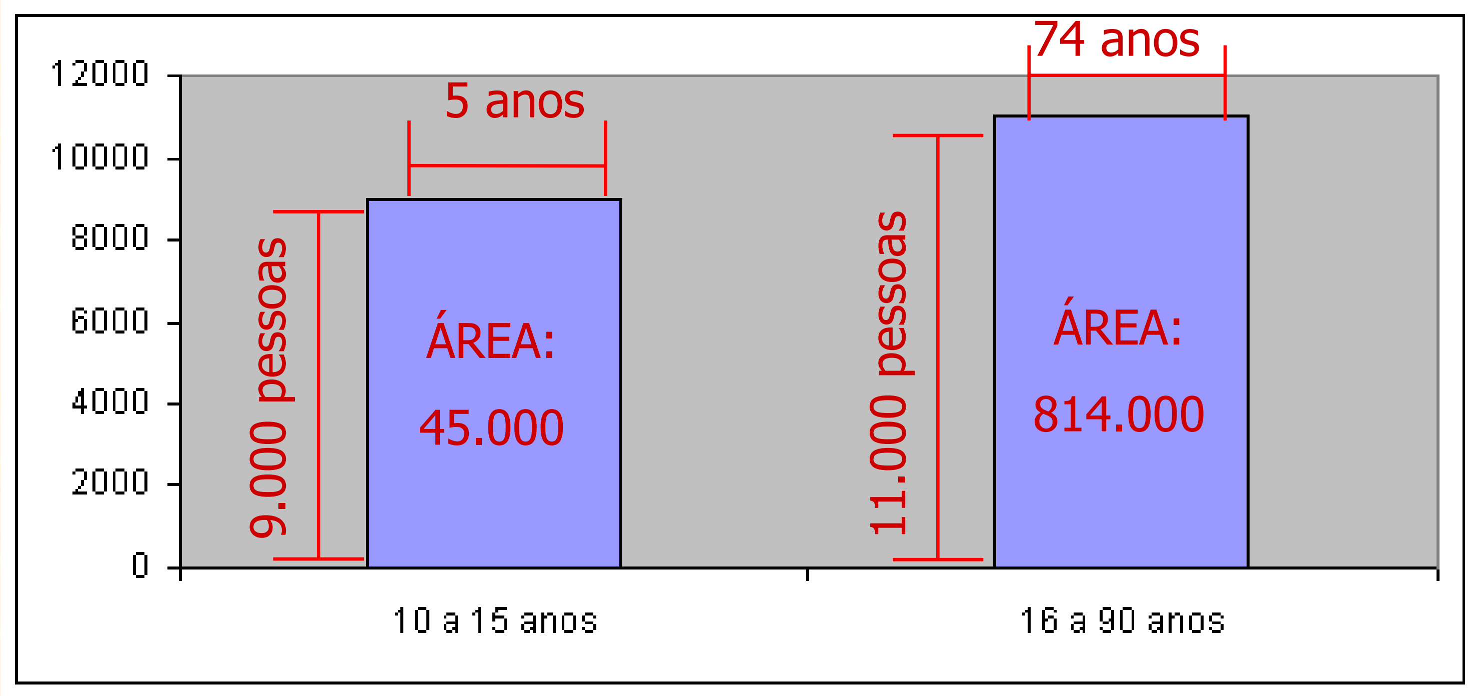


III – Gráficos

* Observações e cuidados

b) Gráficos de Barras, de Setores e Pictóricos: ÁREA

Exemplo: nº de pessoas jogando videogame num evento

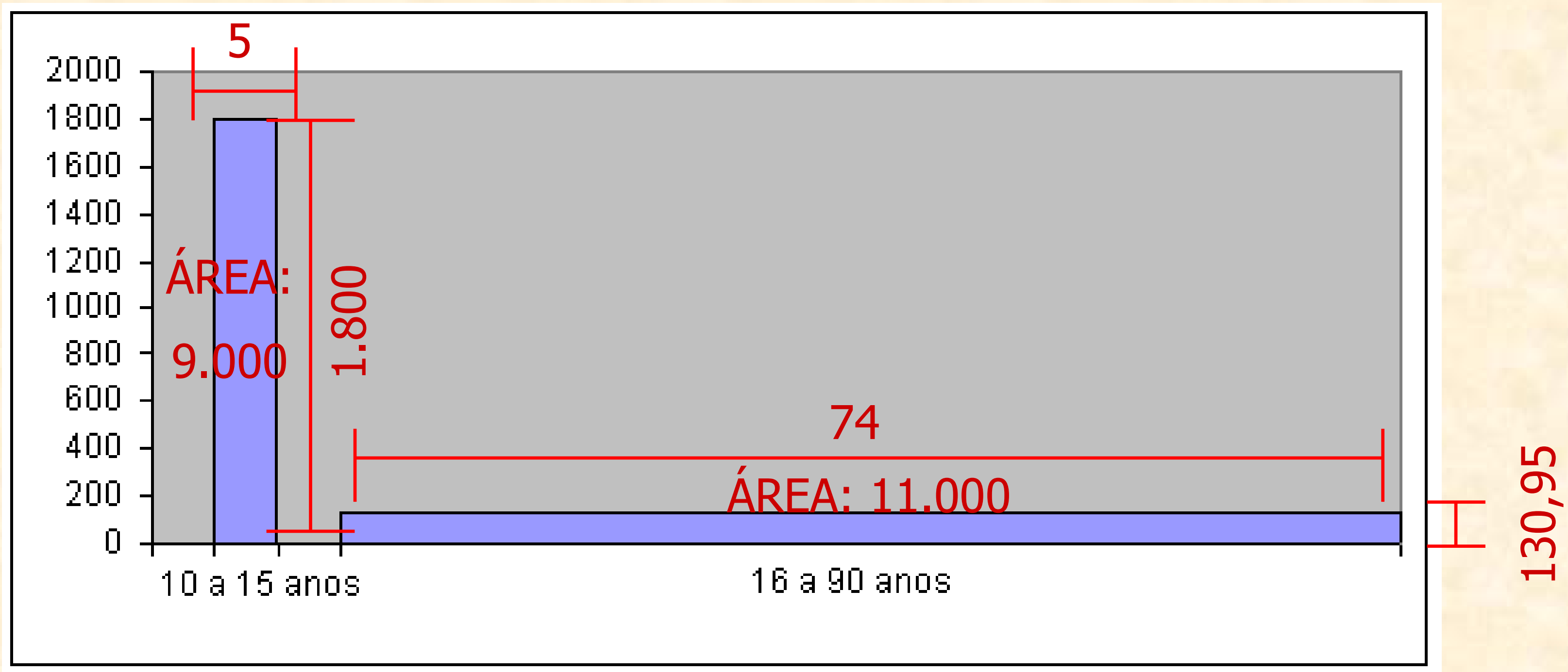


III – Gráficos

* Observações e cuidados

b) Gráficos de Barras, de Setores e Pictóricos: ÁREA

Exemplo: nº de pessoas jogando videogame num evento

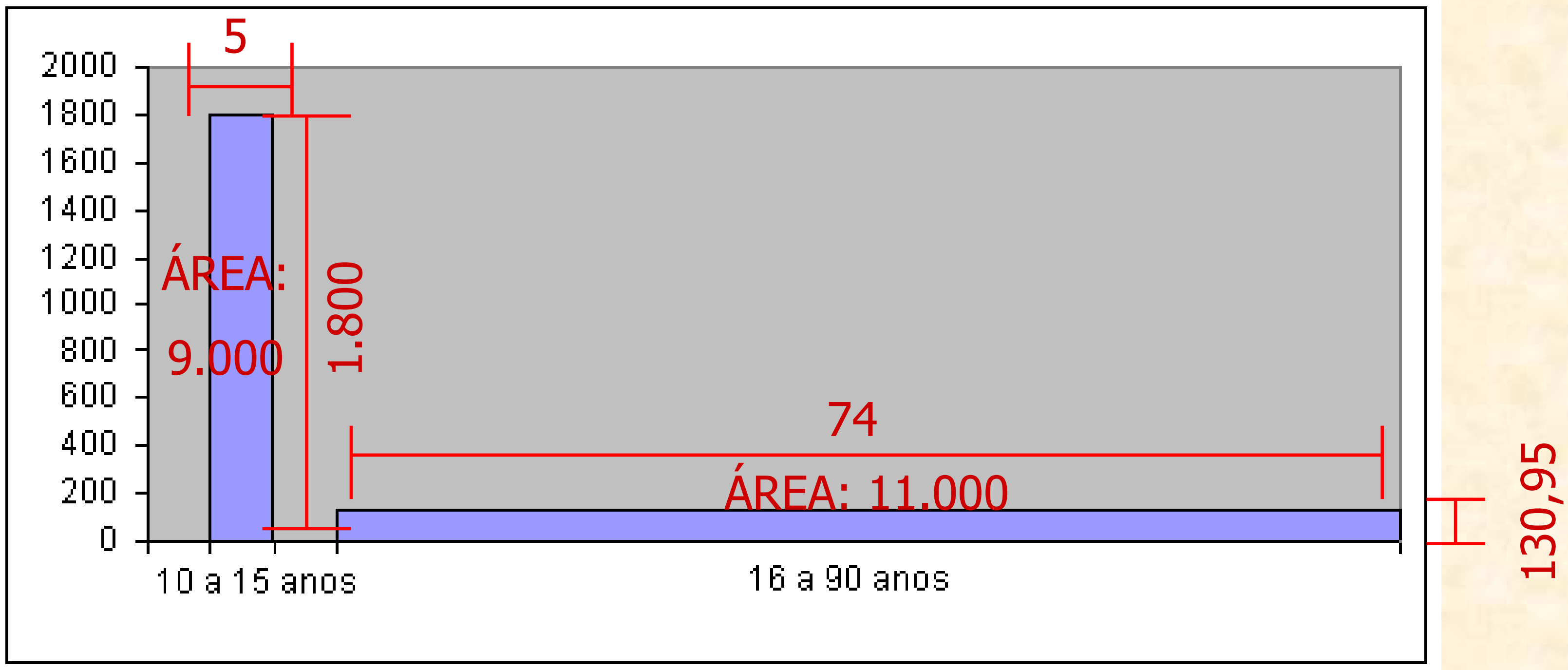


III – Gráficos

* Observações e cuidados

b) Gráficos de Barras, de Setores e Pictóricos: ÁREA

Obs: esse tipo de gráfico se chama **Histograma**.



III – Gráficos

* Observações e cuidados

b) Gráficos de Barras, de Setores e Pictóricos: ÁREA

Exemplo: evolução da venda de aparelhos de TV

III – Gráficos

* Observações e cuidados

b) Gráficos de Barras, de Setores e Pictóricos: ÁREA

Exemplo: evolução da venda de aparelhos de TV

IV - ATENÇÃO

trocar
mudar
voltar
repetir

continuações
remakes
reprises

reserva
substituto
suplente

e Guarani. O Azulão, cada vez mais em decadência, trocou de técnico três vezes na série B. O time começou com Ademir Fonseca, mudou para Márcio Goiano e no final do primeiro turno foi comandado por Vadão. Já o Bragantino, que conta com o técnico Marcelo Veiga desde 2007, vem como um dos piores visitantes, com apenas 14,8% de aproveitamento fora de casa. Situação pior vive o Guarani. O time de Campinas, rebaixado na série A do Brasileiro de 2010, está próximo de cair mais uma vez. Mesmo com a troca do técnico Vilson Taddei (que levou o Bugre de volta à primeira divisão no Paulistão) por Giba, o Guarani não conseguiu melhorar na tabela. Para piorar, o time perdeu seu artilheiro, o atacante Fernandão, que foi para o Palmeiras.

Referências e Sugestões

Livros:

ESTATÍSTICA BÁSICA
Pedro Morettin e Wilton Bussab
Ed. Saraiva

FUNDAMENTOS DE MATEMÁTICA ELEMENTAR
VOLUME 5 – PROBABILIDADES
Samuel Hazzan
VOLUME 11 – ESTATÍSTICA
Gelson Iezzi, Samuel Hazzan, David Degenszajn
Ed. Atual

Referências e Sugestões

Sites: (para quem entende bem inglês, sugiro a leitura da versão anglófona das páginas abaixo)

http://pt.wikipedia.org/wiki/Estatística_descritiva

http://pt.wikipedia.org/wiki/Covariância

http://pt.wikipedia.org/wiki/Coeficiente_de_correlação_de_Pearson

http://pt.wikipedia.org/wiki/Obliquidade

http://pt.wikipedia.org/wiki/Distribuição_de_Bernoulli

http://pt.wikipedia.org/wiki/Distribuição_binomial

http://pt.wikipedia.org/wiki/Distribuição_hipergeométrica

http://en.wikipedia.org/wiki/Charts