Sumário

Capítulo 1: Introdução à Estatística

Definição de estatística

É uma ciência que envolve um corpo de técnicas e uma metodologia desenvolvida para a coleta, a tabulação, a classificação e simplificação de dados, para tornar esses dados melhor apresentáveis, para análise e a interpretação dos mesmos para a tomada de decisões.

Conceitos

População: É um conjunto de elementos com características iguais ou parecidas, agrupadas em conjuntos denominados como população ou universo.

Amostra: É um subconjunto retirado da população com o objetivo de ser analisado, obtendo assim informações dessa amostra, para poder ser generalizado para a população.

Dados experimentais: São dados obtidos de amostras de uma população composta de variáveis.

Censo: É a análise de todos os elementos de uma população.

Observação: Se a população é pequena, muitas vezes é mais indicado um censo do que uma retirada de uma amostra. Na grande maioria das vezes, quando vamos aplicar a estatística na prática, trabalhamos com populações com um grande número de elementos, portanto se torna inviável o estudo de cada elemento da população, pois seria trabalhoso e teríamos um custo alto e em certos casos o estudo de cada elemento da população também é impossível, que seria o caso de determinarmos a média de quilometragem que um determinado pneu é capaz de rodar, pois se testarmos todos os pneus, estaríamos destruindo todos os pneus, portanto trabalhamos tomando uma amostra da população.

Podemos dizer que a estatística se divide em três ramos.

Estatística Descritiva: Trata da coleta, classificação, organização, tabulação, do resumo e, em geral, da simplificação de informações que podem ser muito complexas, fazendo uso de parâmetros estatísticos que resumem o comportamento dos dados.

Exemplo: Taxa de desemprego, custo de vida, quilometragem média por litro de combustível, médias das idades de um grupo de pessoas, etc.

Estatística Probabilística: Utilizada em situações que envolvem o acaso.

Exemplo: Jogos de dados, jogos de cartas, loterias, etc.

O acaso é o que não podemos prever com certeza, é algo que surge ou acontece sem motivo ou explicação aparente.

Estatística Inferencial: Consiste da aplicação de um corpo de técnicas e metodologias para mensurar os resultados obtidos de uma amostra para toda população.

Exemplo: Pesquisa de intenção de votos para presidente de um país.

Capítulo 2: Estatística Descritiva

2.1 Séries estatísticas.

Definição: Uma série estatística nada mais é do que uma tabela, que relaciona um conjunto de dados com uma especificidade qualquer, essa especificidade que me refiro aqui pode ser o tempo, o espaço, ou uma categoria específica, temos assim uma classificação para as séries

As séries estatísticas podem ser classificadas da seguinte forma:

- Séries históricas, cronológicas, temporal ou marcha: representam os dados em função do tempo.
- Séries geográficas, territoriais ou de localização: representam os dados em função da localidade
- Séries específicas ou categóricas: representam os dados em função da especificação ou categoria.

Observação: Entre todas as séries a mais importante é a série temporal, pois o estudo dessa série possibilita fazer estimativas futuras, identificar tendências, e outras análises importantes, utilizando os dados do passado.

Exemplo 2.1

Série Geográfica

País	*PIB em 2007 (em bilhões de dólares)*
China	3380
Brasil	1310
India	800

Série temporal

Ano	*PIB do Brasil (em bilhões de dólares)*
2005	796
2006	956
2007	1310

2.2 Tabulação de Dados

Conceito: Nesta seção veremos técnicas empregadas pela estatística para simplificar e resumir dados em uma tabela, facilitando assim a leitura e interpretação dos mesmos.

2.2.1 Tabela primitiva

Definição: É uma coleção de dados sem nenhum tratamento de organização.

Exemplo 2.2: Suponha que desejamos estimar a média de idade dos estudantes do turno noturno do ensino médio de uma determinada escola e para isso coletamos vinte e uma idades desses estudantes conforme dados abaixo.

18	19	20	18	18	17	17	22	25	24
23	27	21	28	19	20	21	21	23	23
23									

Os dados acima foram coletados e anotados sem nenhuma preocupação organizacional, dizemos então que esta é uma tabela primitiva, pois está sem nenhum tratamento organizacional. Os dados estão desordenados dificultando assim uma análise.

2.2.2 Rol

Definição: É um conjunto de dados numéricos ordenados em ordem crescente.

Exemplo 2.3: Colocando os dados da tabela primitiva acima no rol temos:

17	17	18	18	18	19	19	20	20	21
21	21	22	23	23	23	23	24	25	27
28									

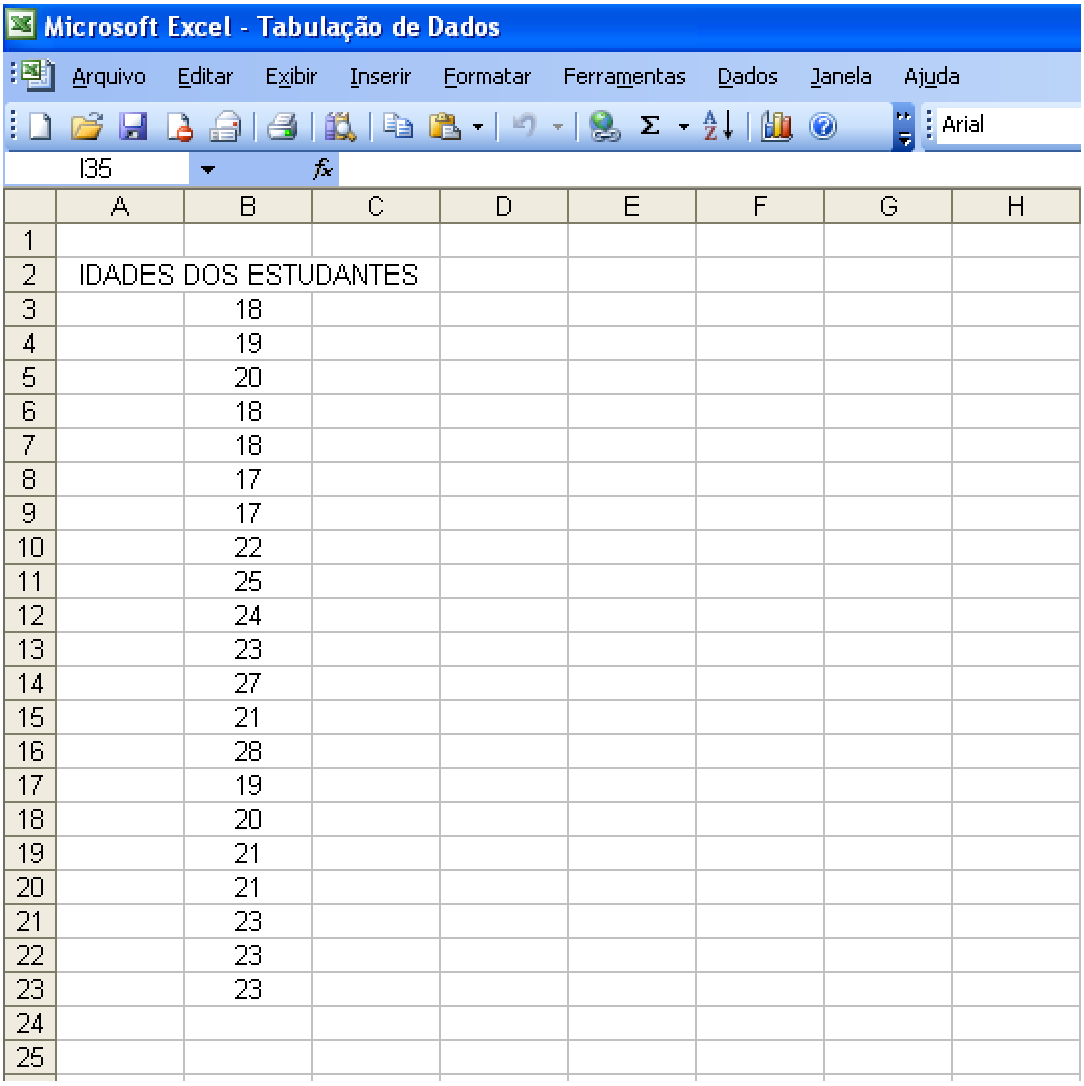
Podemos ver que com esta organização já fica mais fácil a observação de algumas informações tais como o menor dado que é 17, o maior que é 28, quantas vezes se repetem o dado 18, que são três vezes, etc.

Esta organização poderá ser feita no excel, entrando com os dados em uma coluna ou linha e depois clicamos no menu dados / classificar. Veja a seqüência abaixo e confira o arquivo Tabulação de Dados planilha ROL.

Observação: Usamos o Excel 2003 e o 2007 em nossos exemplos.

Entramos com os dados conforme figura 2.1.a abaixo. (Excel 2003)

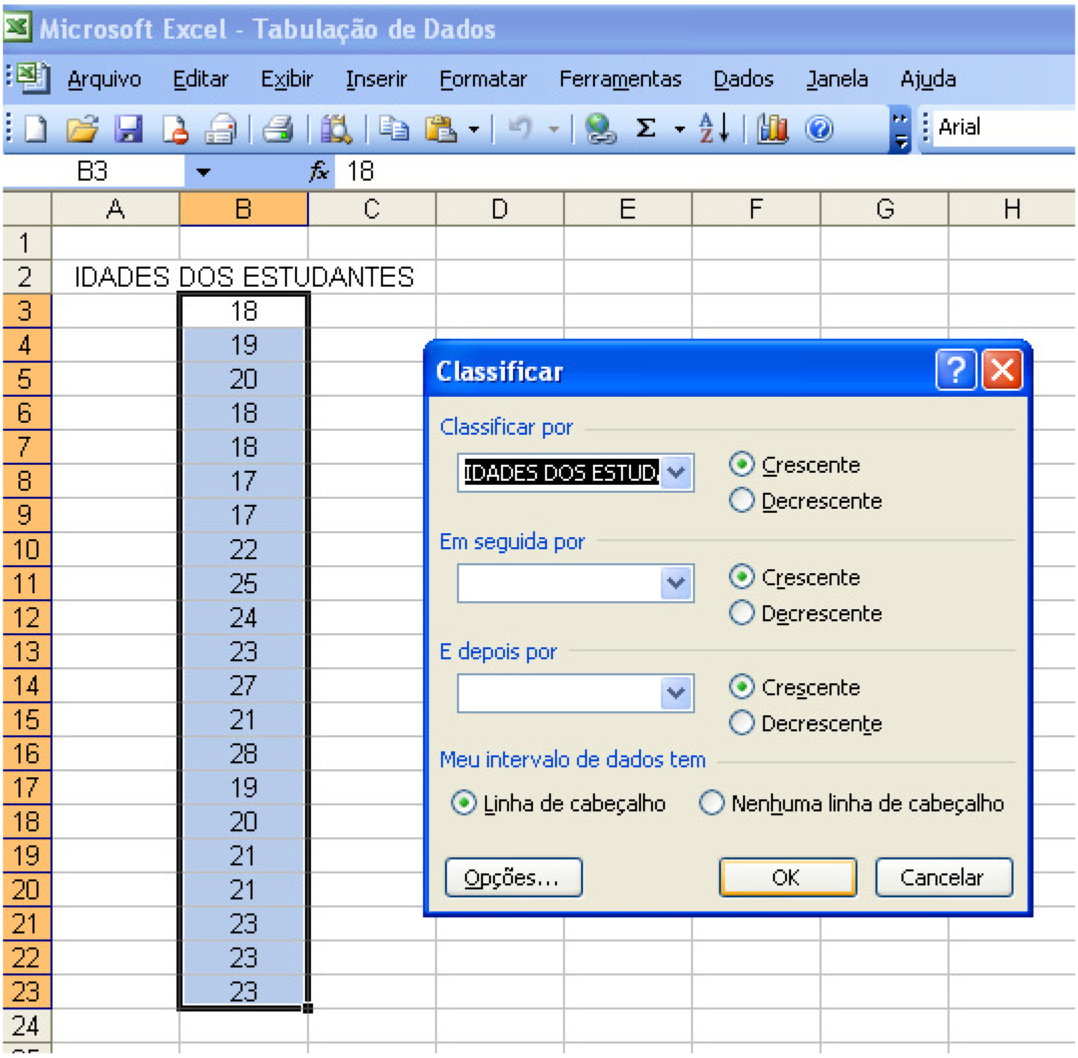
Figura 2.1.a



	A	B	C	D	E	F	G	H
1								
2	IDADES DOS ESTUDANTES							
3		18						
4		19						
5		20						
6		18						
7		18						
8		17						
9		17						
10		22						
11		25						
12		24						
13		23						
14		27						
15		21						
16		28						
17		19						
18		20						
19		21						
20		21						
21		23						
22		23						
23		23						
24								
25								

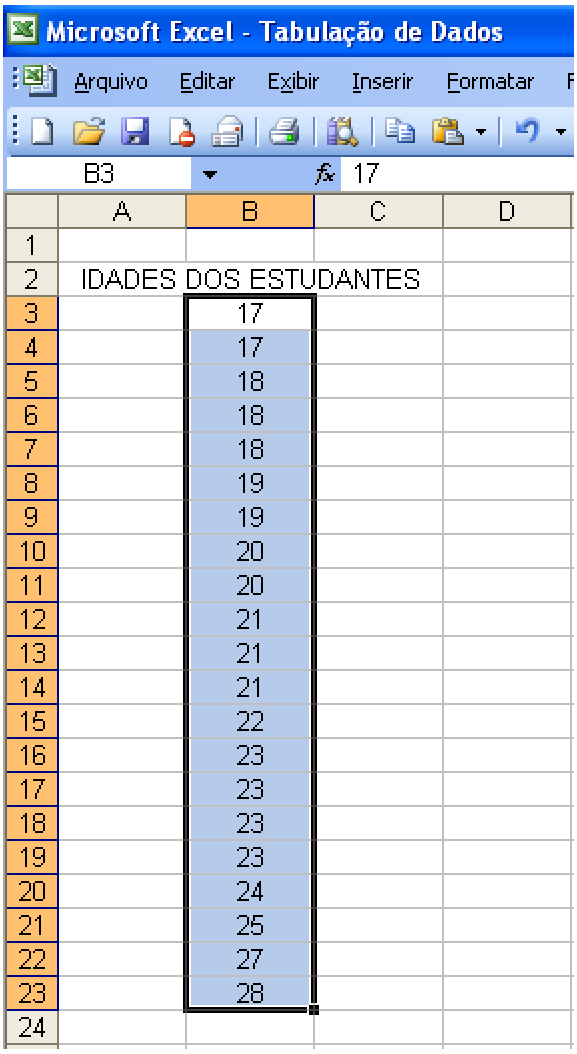
Depois selecionamos os dados, clicamos em dados na barra de ferramentas e escolhemos classificar, aparecerá a seguinte caixa de diálogo conforme figura 2.2.a abaixo.

Figura 2.2.a



Clicando em OK, teremos o rol das idades conforme figura 2.3.a abaixo.

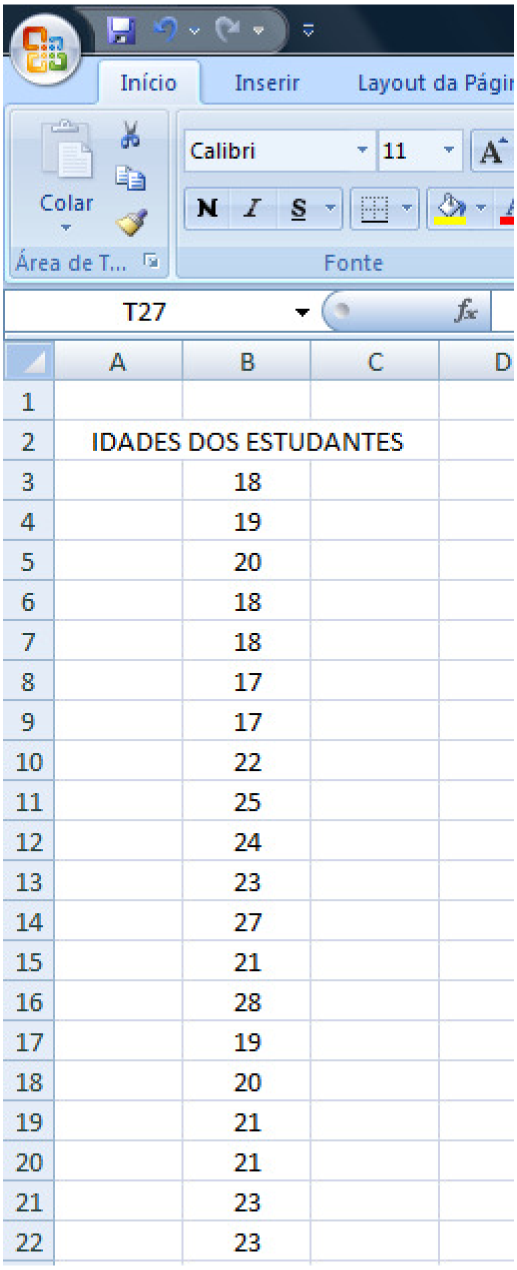
Figura 2.3.a



Este exemplo no Excel 2007 se resolve da seguinte forma:

Entramos com os dados conforme figura 2.1.b abaixo. (Excel 2007)

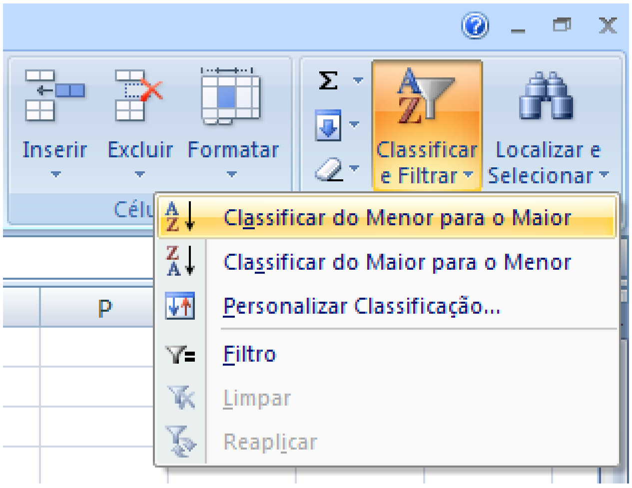
Figura 2.1.b



	A	B	C	D
1				
2		IDADES DOS ESTUDANTES		
3		18		
4		19		
5		20		
6		18		
7		18		
8		17		
9		17		
10		22		
11		25		
12		24		
13		23		
14		27		
15		21		
16		28		
17		19		
18		20		
19		21		
20		21		
21		23		
22		23		

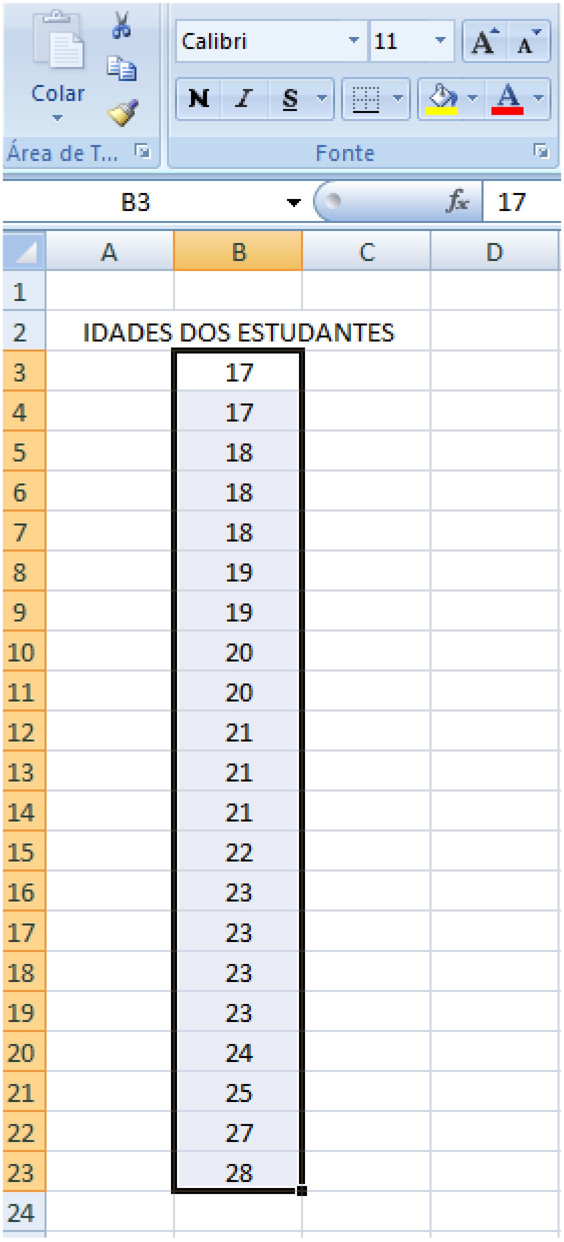
Depois selecionamos os dados, clicamos na barra de ferramentas em classificar e filtrar e selecionamos classificar do menor para o maior conforme caixa de diálogo da figura 2.2.b abaixo.

Figura 2.2.b



Teremos assim o rol das idades conforme figura 2.3.b abaixo.

Figura 2.3.b



	A	B	C	D
1				
2	IDADES DOS ESTUDANTES			
3		17		
4		17		
5		18		
6		18		
7		18		
8		19		
9		19		
10		20		
11		20		
12		21		
13		21		
14		21		
15		22		
16		23		
17		23		
18		23		
19		23		
20		24		
21		25		
22		27		
23		28		
24				

2.2.3 Distribuição de freqüência

Definição: É uma tabela em que temos os dados em uma coluna, e na outra coluna a freqüência que cada dado se repete.

Exemplo 2.4 Podemos melhorar a organização desses dados, dispondo-os de forma que apareça o número de vezes em que se repetem cada dado na tabela. A esta organização nós chamamos de distribuição de freqüência.

Distribuição de Freqüência

Idade	Freqüência
17	2
18	3
19	2
20	2
21	3
22	1
23	4
24	1
25	1
27	1
28	1
$\sum$	21

Na representação em distribuição de freqüência é importante colocar as freqüências acumuladas absolutas e a as freqüências relativas e relativas acumuladas, pois essas informações na distribuição facilitará a análise dos dados, vejamos então a distribuição acima com essas informações.

Distribuição de Freqüência

Idade	Freqüência (F_i)	Freqüência% (F_i%)	Freqüência Acumulada (Fac_i)	Freqüência Acumulada % (Fac_i%)
17	2	9,52	2	9,52
18	3	14,29	5	23,81
19	2	9,52	7	33,33
20	2	9,52	9	42,85
21	3	14,29	12	57,14
22	1	4,76	13	61,90
23	4	19,05	17	80,96
24	1	4,76	18	85,72
25	1	4,76	19	90,48
27	1	4,76	20	95,24
28	1	4,76	21	100
$\sum$	21	100		

Vamos adotar aqui uma notação simples para as freqüências: Para freqüência absoluta da classe i, vamos usar F_i, para freqüência relativa vamos usar $F_i\%$, para a freqüência absoluta acumulada, vamos usar Fac_i, e para relativa acumulada $Fac_i\%$.

A coluna da freqüência acumulada nos dá uma informação importante, por exemplo, olhando para coluna da freqüência acumulada, vemos facilmente que 12 dados são menores ou iguais a 21, olhando para coluna da freqüência acumulada relativa (%) vemos facilmente que isso corresponde a 57,14% dos dados, essa informações estão destacadas na tabela acima em verde.

Podemos construir essa distribuição de freqüência, assim como um gráfico de barras que representa a freqüência de cada dado, que chamamos de histograma. Para isso clicamos na barra de ferramentas do Excel 2003 em *ferramentas* e depois clicamos em *análise de dados* e depois clicamos em *histograma*. Veja essa sequência abaixo nas figuras 2.4 e 2.5:

Figura 2.4

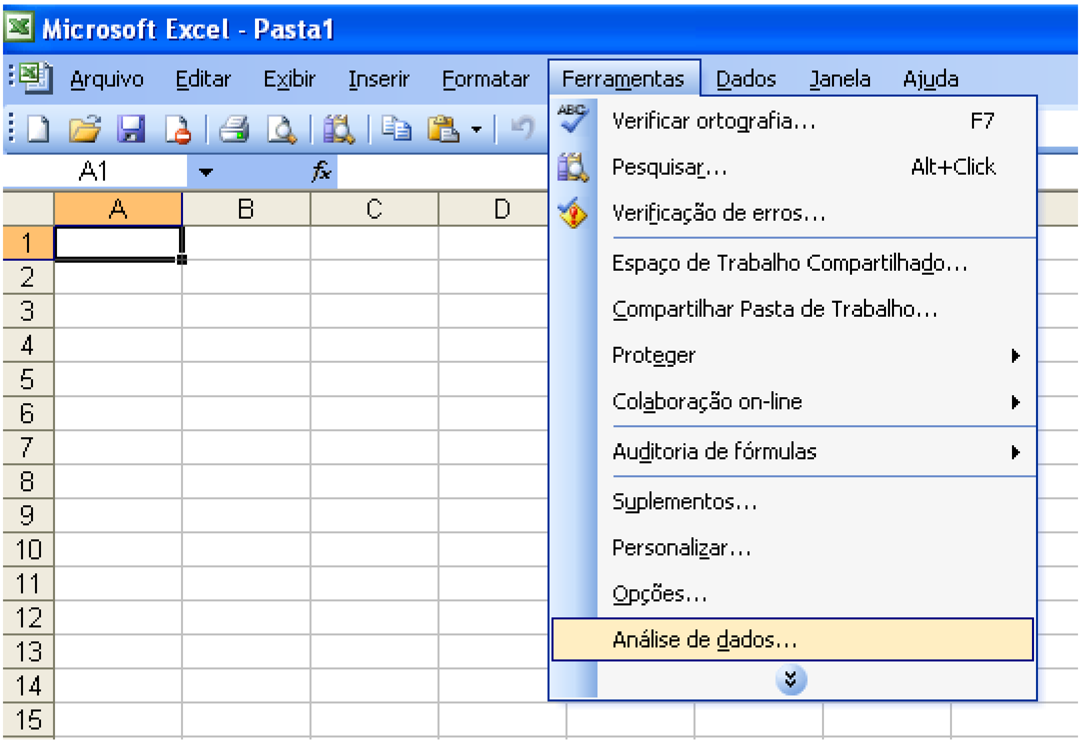
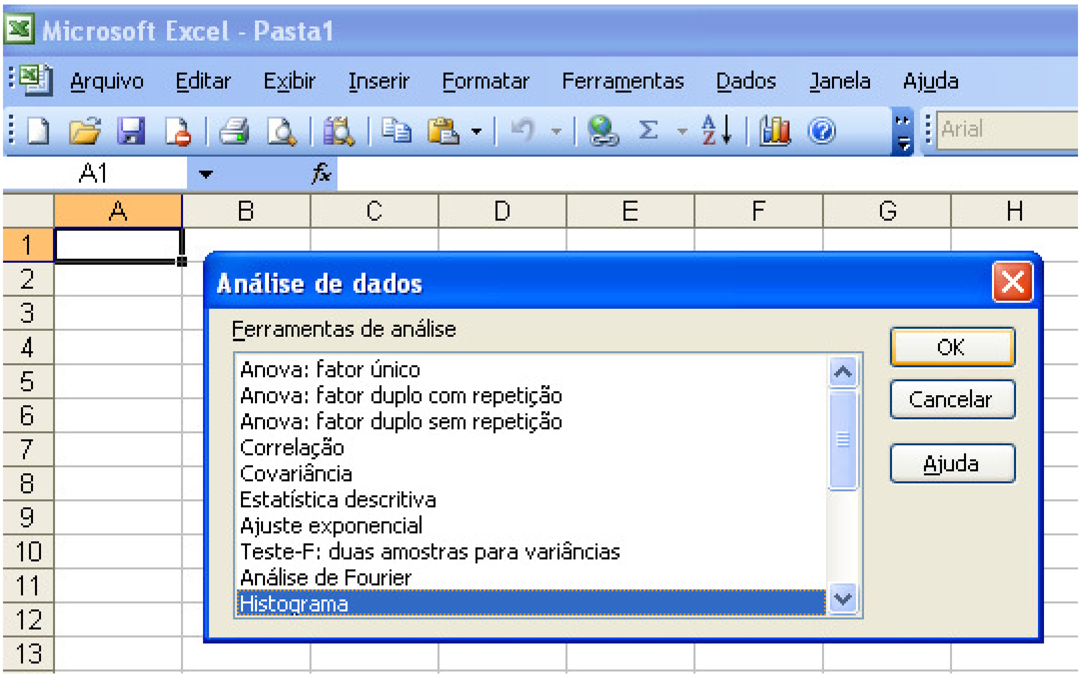


Figura 2.5



Caso não apareça a opção *análise de dados,* clique na barra de ferramentas em *ferramentas* depois clique em suplementos e selecione *ferramentas de análise* clicando em seguida em ok, para instalar as ferramentas de análise. Caso não instale, você terá que inserir o cd de instalação do Ms Office para instalar essas ferramentas. Veja essa sequência abaixo nas figuras 2.6 e 2.7.

Figura 2.6

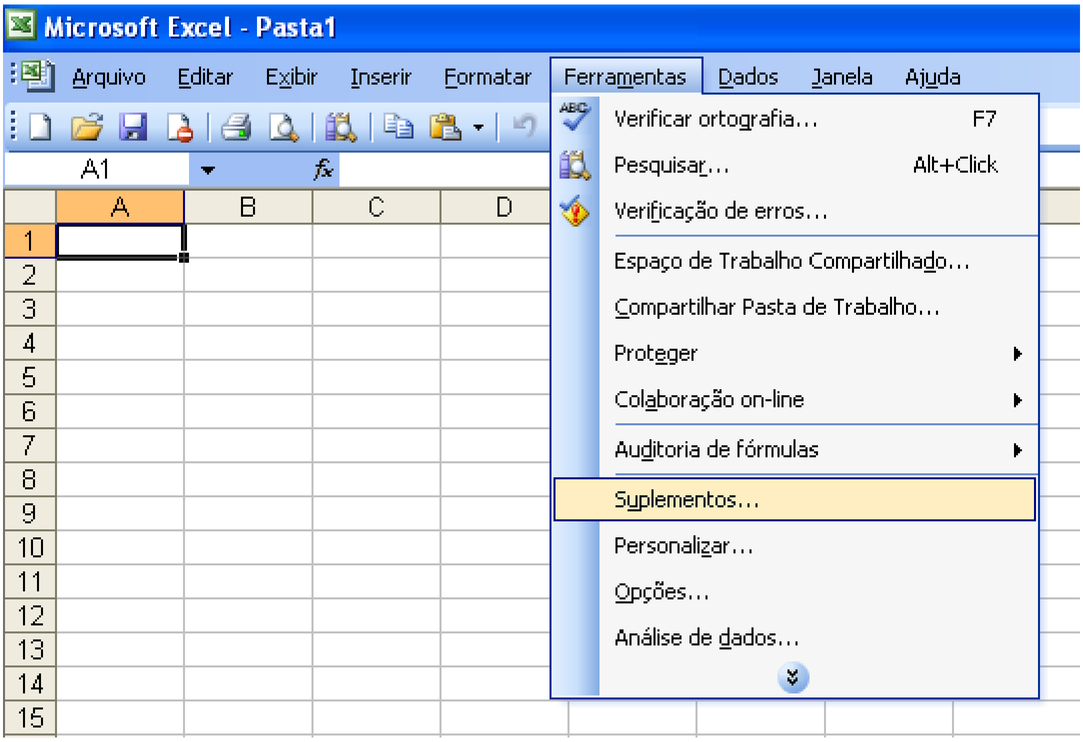
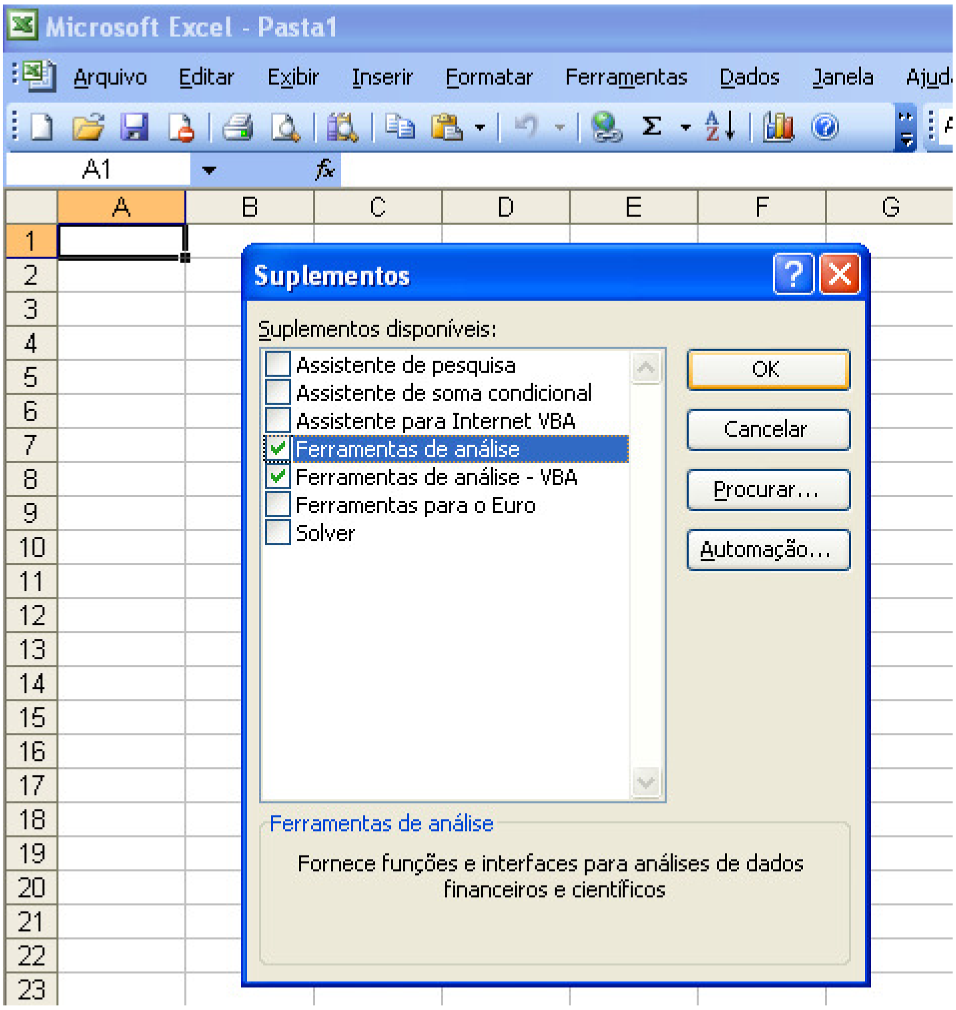
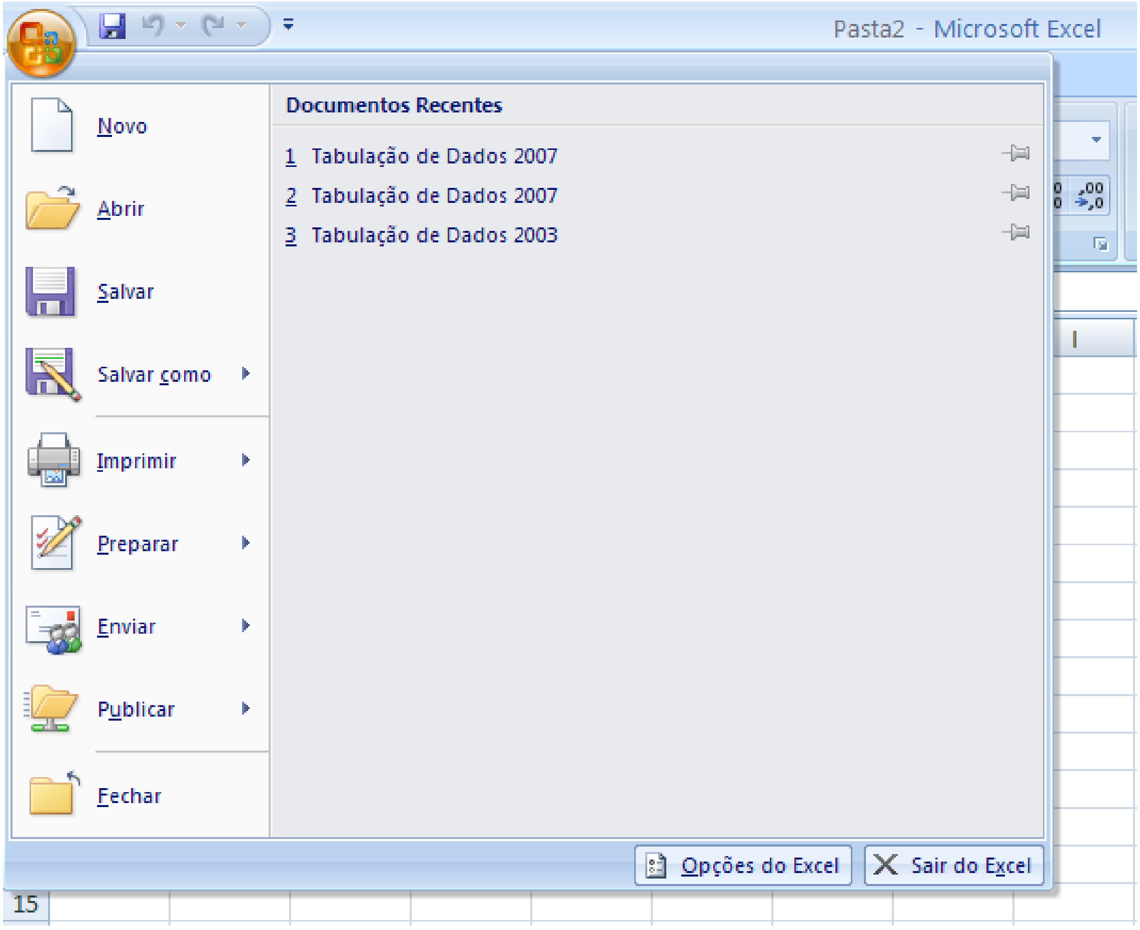

Microsoft Excel - Pasta1
Arquivo
Editar
Exibir
Inserir
Formatar
Ferramentas
Dados
Janela
Ajuda
A1
Verificar ortografia... F7
Pesquisar... Alt+Click
Verificação de erros...
Espaço de Trabalho Compartilhado...
Compartilhar Pasta de Trabalho...
Proteger
Colaboração on-line
Auditoria de fórmulas
Suplementos...
Personalizar...
Opções...
Análise de dados...

Figura 2.7



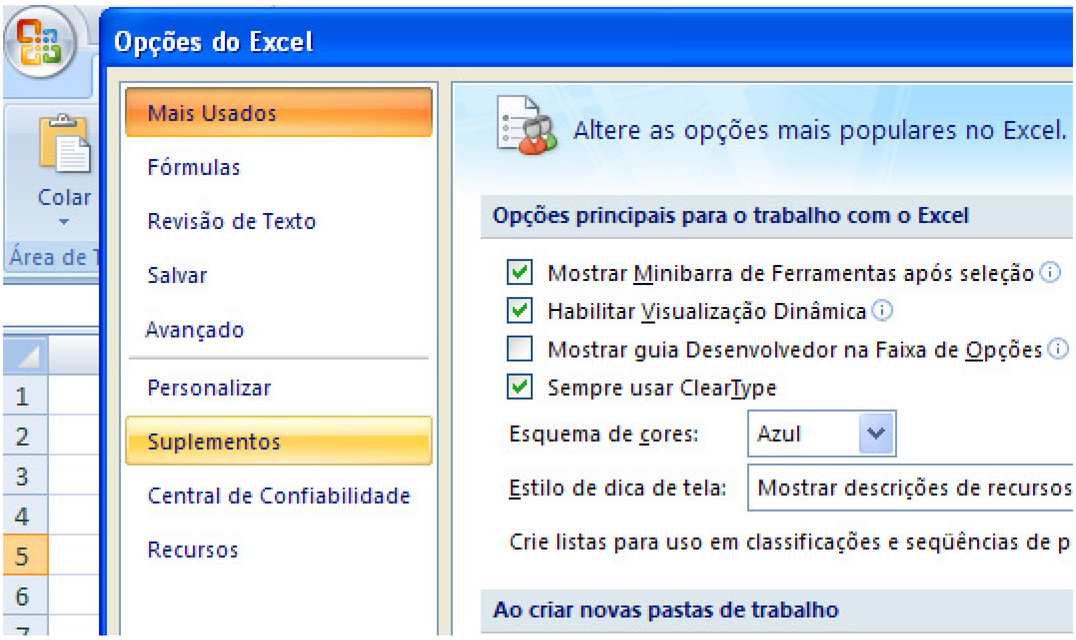
No Excel 2007, para instalar as ferramentas de análise nós clicamos no botão do Office, depois em opções do Excel. Ver figura 2.8 abaixo:

Figura 2.8



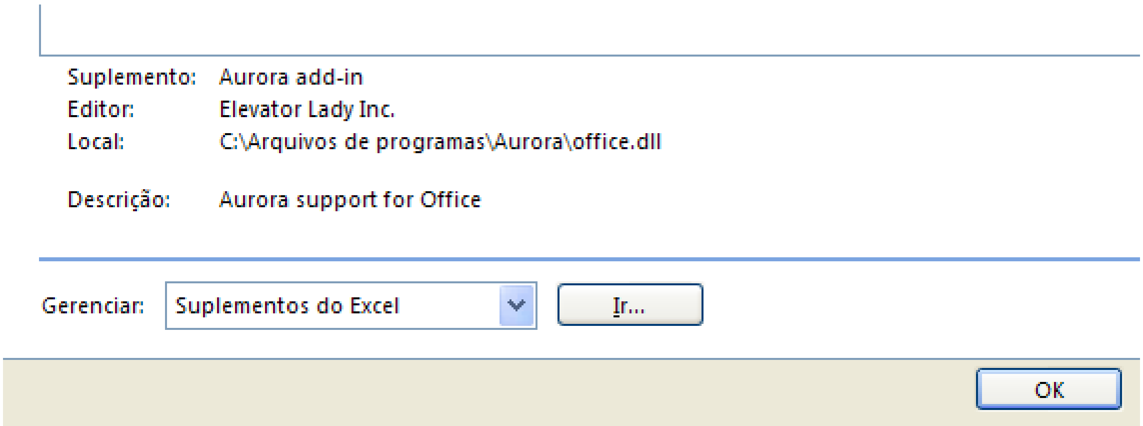
Depois clicamos em suplementos, ver figura 2.9 abaixo:

Figura 2.9



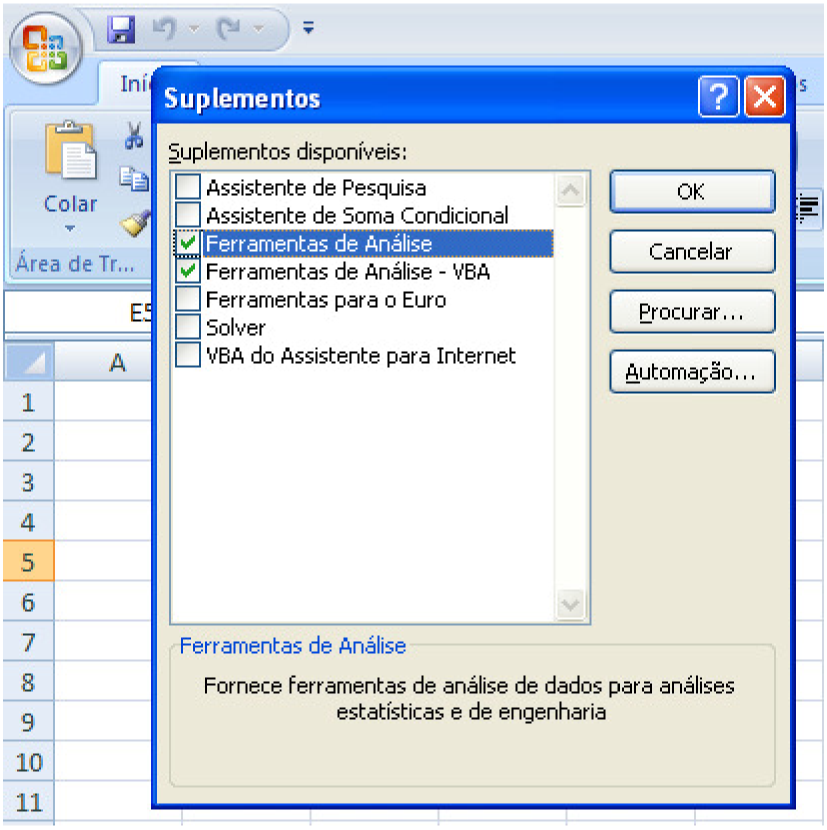
Em gerenciar selecionamos suplementos do Excel, cliclamos em ir. Ver figura 2.10 abaixo:

Figura 2.10



depois selecionamos ferramentas de análise e clicamos em ok. Ver figura 2.11 abaixo:

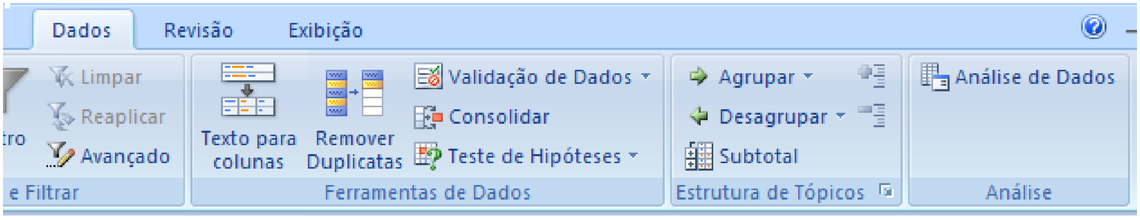
Figura 2.11



Quando a instalação estiver terminada você verá na barra de ferramentas, clicando em dados, a opção análise de dados.

Caso não apareça na barra de ferramentas em dados a opção análise de dados, você deve inserir o cd do Office para instalar as ferramentas completa do Office 2007. Veja na figura 2.12 abaixo a barra de ferramentas do Excel 2007 como aparecerá a opção análise de dados.

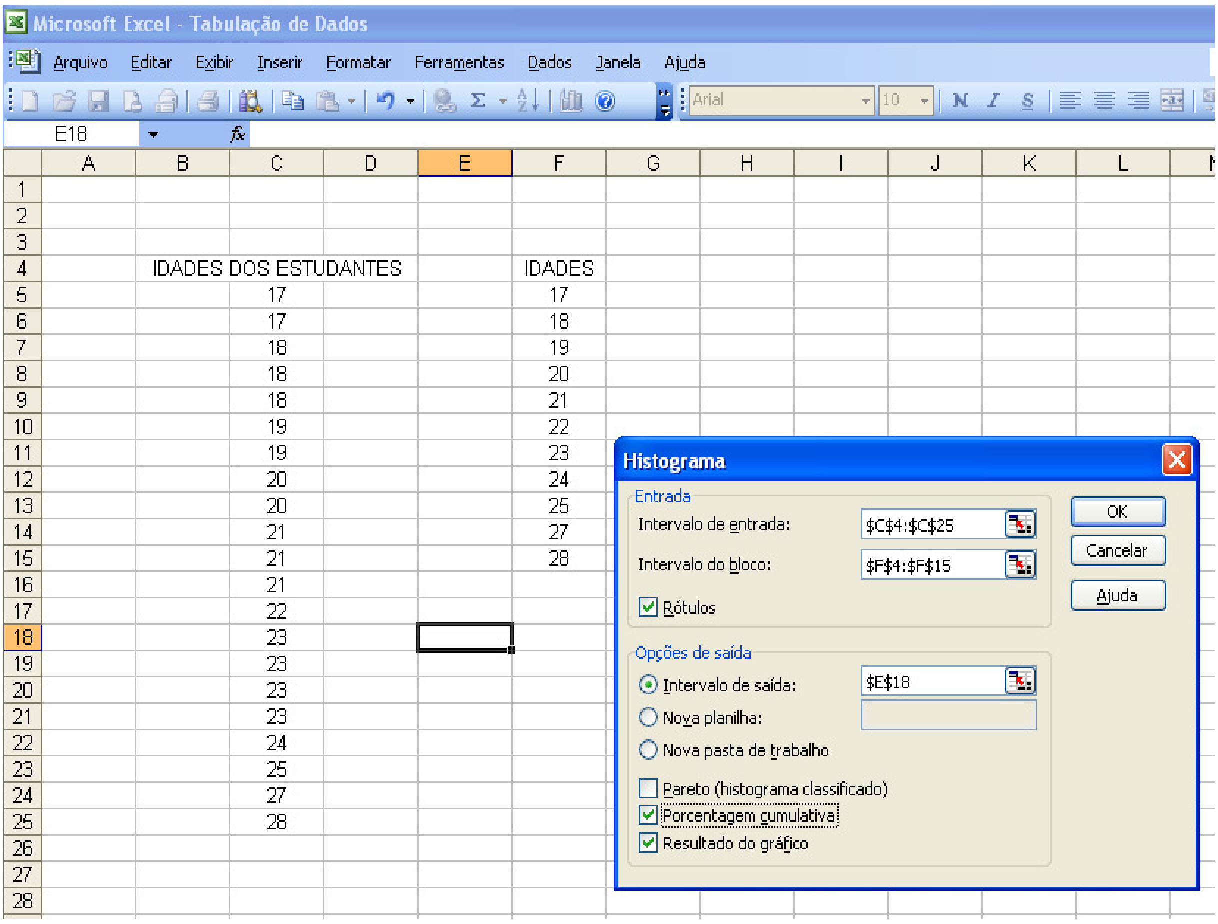
Figura 2.13



Vejamos como construir essa distribuição de freqüência no Excel 2003.

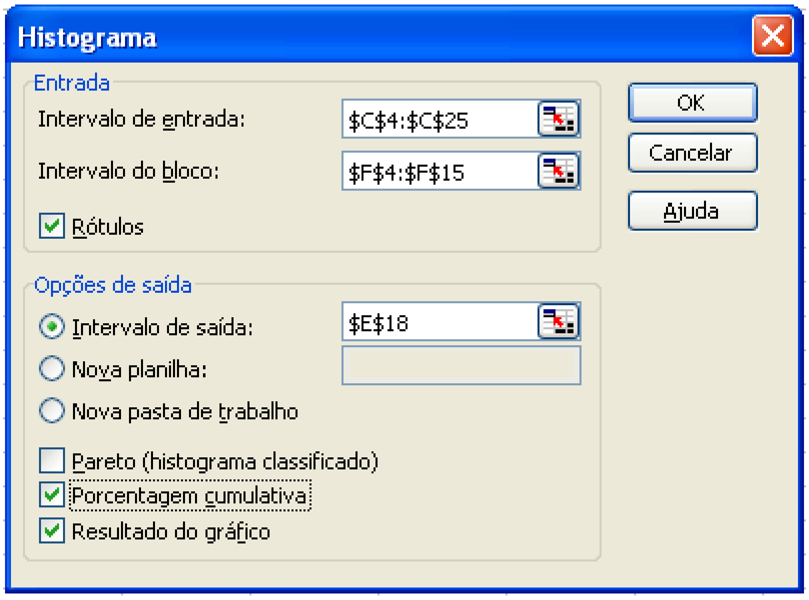
No arquivo Tabulação de Dados 2003, na planilha distribuição de freqüência, criamos a coluna com os dados das idades, e fomos em *ferramentas / analise de dados* e selecionamos a opção *histograma* conforme figura 211.a abaixo:

Figura 2.11.a



Na caixa de diálogo como intervalo de entrada selecionamos da célula C4 a célula C25 e o intervalo de bloco da célula F4 até a célula F15, selecionamos o rótulo, pois incluímos as células C4 e F4, onde contém os títulos/rótulos *idades dos estudantes* e *idades* respectivamente, pois assim aparecerá no gráfico a referência *idades* contida na célula F4. No intervalo de saída, optamos para que o gráfico seja construído na mesma planilha e escolhemos que fosse construído a partir da célula E18, escolhemos para esboçar no gráfico a *porcentagem cumulativa* e o *resultado gráfico*, conforme figura 2.12.a abaixo.

Figura 2.12.a



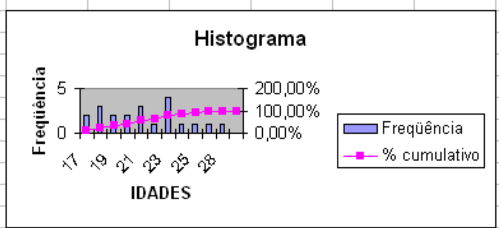
Clicando em ok, teremos a distribuição de freqüência com a freqüência relativa acumulada conforme figura 2.13.a abaixo.

Figura 2.13.a

IDADES DOS ESTUDANTES
17
17
18
18
18
19
19
20
20
21
21
21
22
23
23
23
23
24
25
27
28

IDADES
17
18
19
20
21
22
23
24
25
27
28

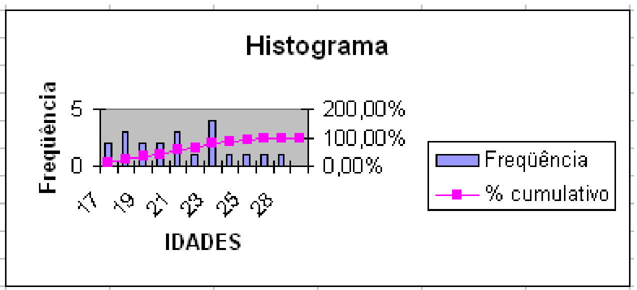
IDADES	*Freqüência*	*% cumulativo*
17	2	9,52%
18	3	23,81%
19	2	33,33%
20	2	42,86%
21	3	57,14%
22	1	61,90%
23	4	80,95%
24	1	85,71%
25	1	90,48%
27	1	95,24%
28	1	100,00%
Mais	0	100,00%



Vejamos com mais detalhes na figura 2.14.a, a tabela e o gráfico que foram construídos:

Figura 2.14.a

IDADES	Freqüência	% cumulativo
17	2	9,52%
18	3	23,81%
19	2	33,33%
20	2	42,86%
21	3	57,14%
22	1	61,90%
23	4	80,95%
24	1	85,71%
25	1	90,48%
27	1	95,24%
28	1	100,00%
Mais	0	100,00%



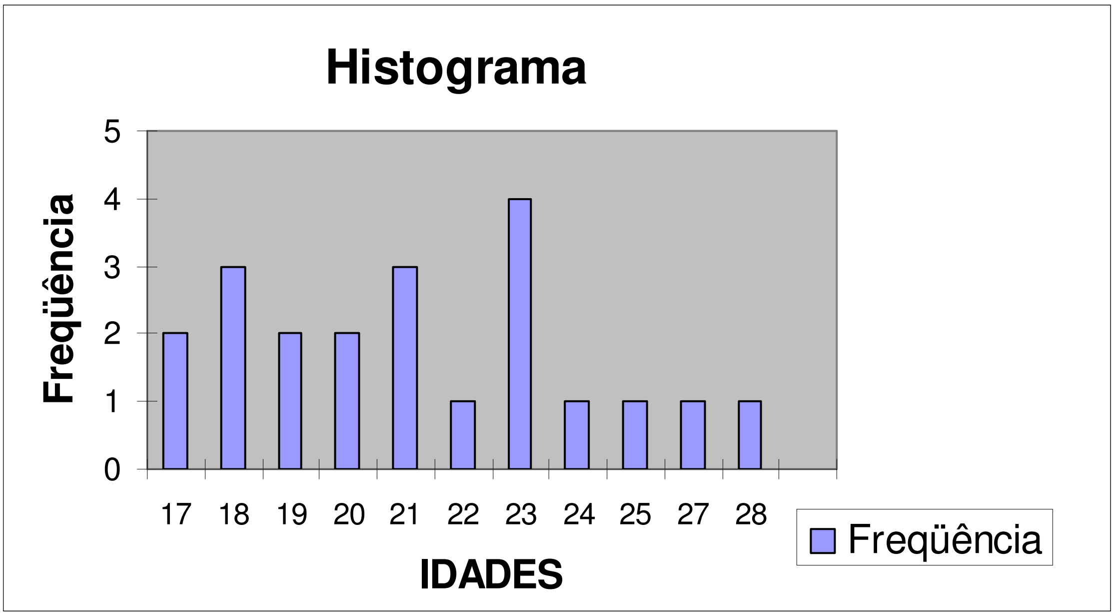
Note que não foi construindo a coluna das freqüências absolutas acumulada e a coluna das freqüências relativas. Podemos retirar na representação gráfica a linha rosa que representa as frequências relativas acumuladas, se assim desejamos, e podemos construir na tabela as colunas com as representações que faltam, ficando o resultado final conforme figura 2.15.a abaixo.

Figura 2.15.a

IDADES	Freqüência	Frequência %	Frequência Acumulada	Frequência Acumulada %
17	2	9,52%	2	9,52%
18	3	14,29%	5	23,81%
19	2	9,52%	7	33,33%
20	2	9,52%	9	42,86%
21	3	14,29%	12	57,14%
22	1	4,76%	13	61,90%
23	4	19,05%	17	80,95%
24	1	4,76%	18	85,71%
25	1	4,76%	19	90,48%
27	1	4,76%	20	95,24%
28	1	4,76%	21	100,00%
Soma	21	100,00%		

E obtemos o histograma conforme figura 2.16.a abaixo.

Figura 2.16.a

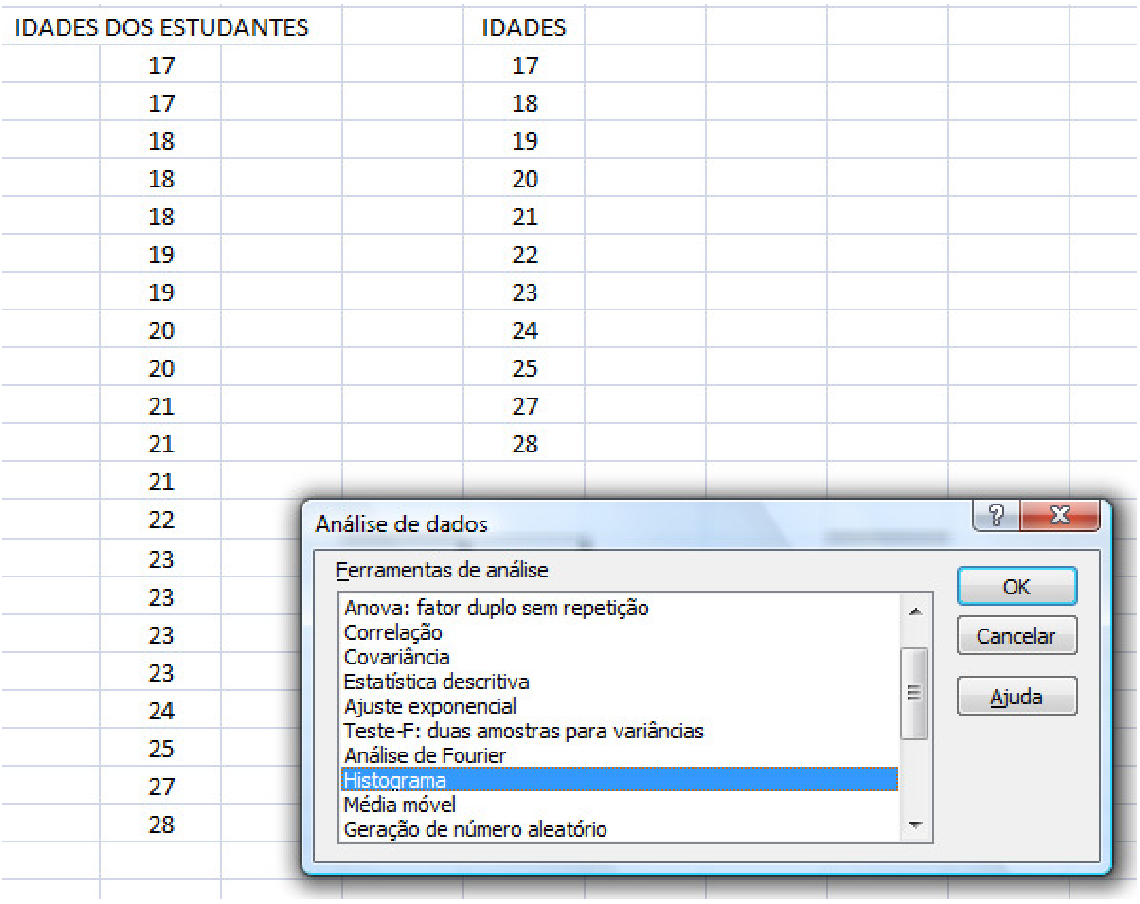


Podemos fazer os ajustes que desejarmos no resultado final apresentado pela função histograma usando conhecimentos básicos de Excel, consulte o arquivo tabulação de dados para poder ver melhor os ajustes feitos e apresentados nas figuras.

Vejamos como construir essa distribuição de freqüência no Excel 2007.

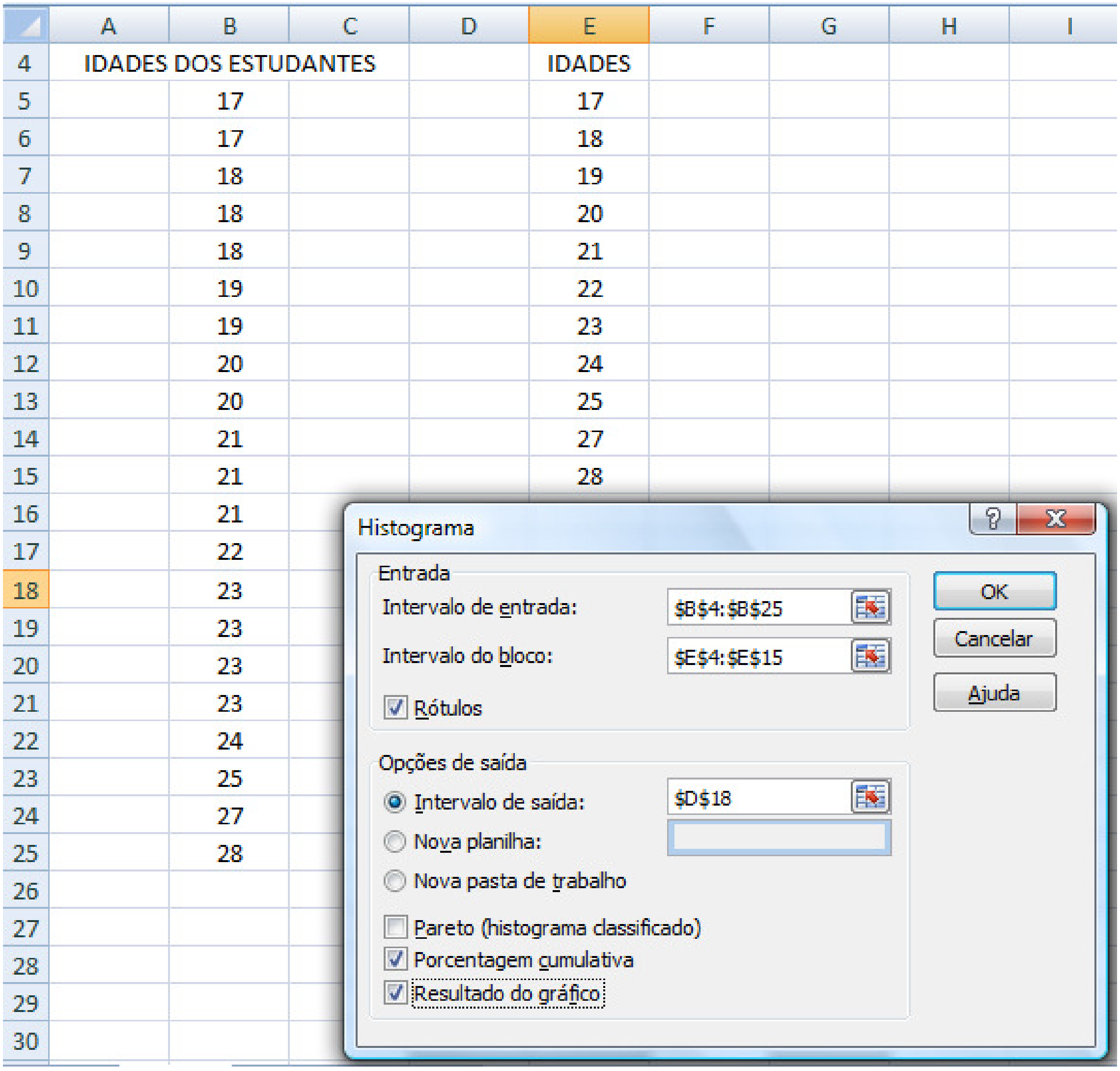
No arquivo Tabulação de Dados 2007, na planilha distribuição de freqüência, criamos a coluna com os dados das idades, e fomos em *dados / analise de dados* e selecionamos a opção *histograma* conforme figura 2.11.b abaixo:

Figura 2.11.b



Na caixa de diálogo como intervalo de entrada selecionamos da célula C4 a célula C25 e o intervalo de bloco da célula F4 até a célula F15, selecionamos o rótulo, pois incluímos as células C4 e F4, onde contém os títulos/rótulos *idades dos estudantes* e *idades* respectivamente, pois assim aparecerá no gráfico a referência *idades* contida na célula F4. No intervalo de saída, optamos para que o gráfico seja construído na mesma planilha e escolhemos que fosse construído a partir da célula E18, escolhemos para esboçar no gráfico a *porcentagem cumulativa* e o *resultado gráfico*, conforme figura 2.12.b abaixo.

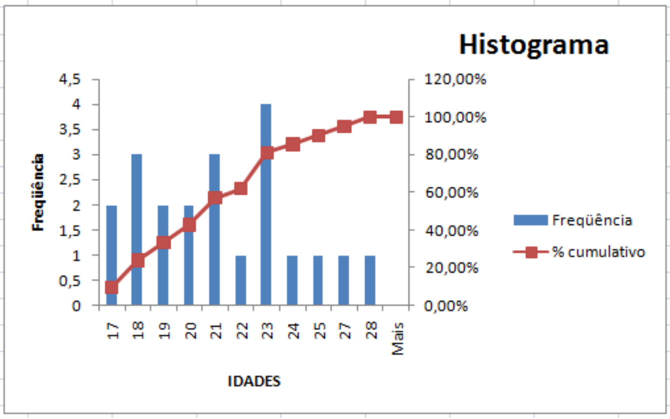
Figura 2.12.b



Clicando em ok, teremos a distribuição de freqüência com a freqüência relativa acumulada conforme figura 2.13.b abaixo.

Figura 2.13.b

IDADES	Freqüência	% cumulativo
17	2	9,52%
18	3	23,81%
19	2	33,33%
20	2	42,86%
21	3	57,14%
22	1	61,90%
23	4	80,95%
24	1	85,71%
25	1	90,48%
27	1	95,24%
28	1	100,00%
Mais	0	100,00%



Os demais procedimentos são iguais aos descritos anteriormente.

Exercícios 2.2.3 (Distribuição de Freqüência)

1) Os dados abaixo se referem aos pesos de um grupo de pessoas:

18	20	21	20	19	18	19	22
19	21	23	19	20	17	17	23
17	18	18	21	17	20	18	17
18	22	17	18	18	21	20	24

Construa a distribuição de freqüência e a sua representação gráfica em um histograma.

2) Agora tomamos as alturas desses 32 em estudantes (em metro) conforme tabela abaixo:

1,64	1,63	1,59	1,91	1,88	1,70	1,56	1,63
1,76	1,75	1,58	1,80	1,82	1,72	1,57	1,78
1,84	1,61	1,63	1,78	1,83	1,71	1,59	1,74
1,57	1,62	1,88	1,79	1,81	1,69	1,61	1,73

Construa a distribuição de frequência e construa o histograma.

A resolução dos exercícios acima encontra-se no arquivo exercícios 2.2.3

2.2.4 Distribuição de frequência em intervalos de classes

Se o rol for muito grande e com dados com poucas repetições torna-se inviável a disposição dos dados em uma distribuição de freqüência, pois teríamos uma tabela muito extensa, (veja como exemplo o exercício 2 da seção anterior) então usamos intervalos de classes, e chamamos este tipo de disposição, de distribuição de freqüência em intervalos de classes. Na prática construímos os intervalos de acordo com nosso interesse, mais adiante veremos uma regra para construção de uma distribuição de freqüência em intervalos de classes, chamada regra de Sturges. Na tabela abaixo o símbolo $\vdash\!\!-$ significa que o intervalo é fechado à esquerda e aberto na direita e o símbolo $\vdash\!\!\dashv$ é fechado nos extremos. Na distribuição de freqüência em intervalos de classe abaixo temos na segunda linha o intervalo que vai de 17 fechado a 19 aberto, olhando para nosso rol do exemplo 2.3 temos 5 dados neste intervalo e assim sucessivamente, temos 7 dados na segunda classe e por fim temos 2 dados na última classe que vai de 26 a 28, intervalo fechado nos extremos.

Observação: Os nossos exemplos se referem aos dados das idades do exemplo 2.2.

Rol

17	17	18	18	18	19	19	20	20	21
21	21	22	23	23	23	23	24	25	27
28									

Exemplo 2.4: Distribuição de frequência em intervalos de classe

i	Idade	Freqüência
1	17 $\vdash\!\!-$ 19	5
2	19 $\vdash\!\!-$ 22	7
3	22 $\vdash\!\!-$ 24	5
4	24 $\vdash\!\!-$ 26	2
5	26 $\vdash\!\!\dashv$ 28	2
	$\sum$	21

2.2.4.1 Componentes de uma distribuição de freqüência em classes

Classe: É o intervalo entre as variáveis estudadas, é denotada pela letra i. O número de classes é denotada pela letra K. No exemplo acima temos cinco classes (K=5). O intervalo 22 $\vdash\!\!-$ 24 está na terceira classe (i=3).

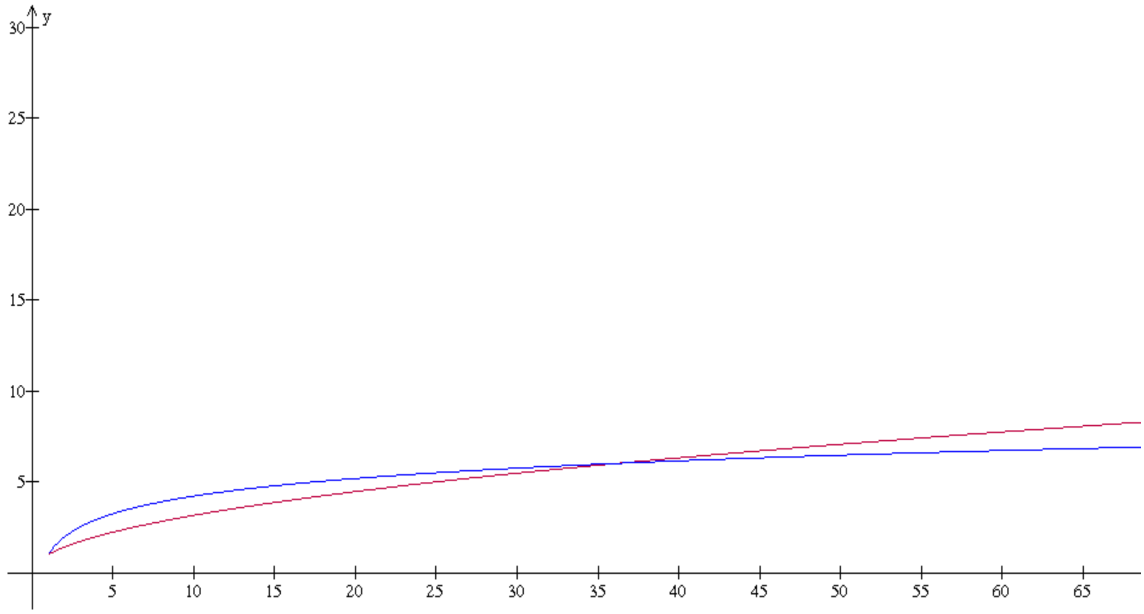
Para determinarmos o número de classes de uma distribuição em classes de freqüência utilizando a regra de Sturges estipulamos que:

Se $n \leq 25$, então K = 5.

Se $n > 25$, então, $k \cong 1 + \log_2 n = 1 + \dfrac{\log n}{\log 2}$. Como $\dfrac{1}{\log 2} \cong 3{,}32$, costumamos usar $K \cong 1 + 3{,}32 \log n$ podemos tomar aqui $K \cong \sqrt{n}$, pois os valores para 1+3,32log$n$ são próximos dos valores de $\sqrt{n}$, porém à medida que os valores de n cresce, temos que os valores de $\sqrt{n}$, cresce bem mais que os valores de 1+3,32log$n$, portanto para valores grandes de n é mais indicado usarmos K ≅ 1 + 3,32log$n$, abaixo vemos na figura 2.7 o gráfico das funções $f(x) = 1 + 3{,}32 \log x$ em azul e

o gráfico da função $g(x) = \sqrt{x}$ em vermelho que mostram graficamente o que acabamos de explicar.
Obs: $\log x$ é o logaritmo de x, na base 10.

Figura 2.7



Por exemplo, se estivermos trabalhando com 1000 dados, ou seja, *n*=1000 teremos pela regra de Sturges K$\cong 1+3{,}32\log 100 \cong 11$ classes, enquanto que se usarmos a raiz quadrada teremos k$\cong \sqrt{1000} \cong 32$ classes temos portanto uma diferença de 21 classes de uma regra para outra.

De um modo geral o número de classes deve ser maior ou igual a 5 e menor ou igual a 25, portanto se tomarmos a regra da raiz como aproximação da regra de Sturges, teremos uma quantidade de dados limitada por aproximadamente 646 dados, pois $\sqrt{646} \cong 25{,}42$, logo teríamos K=25. No caso da regra de Sturges temos como limite de quantidades de dados para termos 25 classes um valor aproximado de 22.668.443 dados, pois $1+3{,}32\log(22668443) \cong 25{,}42$, logo K=25. Quando tivermos trabalhando com quantidades de dados que aplicando a regra de Sturges, o número de classes supere 25 classes, é indicado que se use sempre 25 classes.

Limite de classes

São os extremos de cada classe, o menor valor é o limite inferior (li), e o maior o limite superior (Li). Na distribuição de freqüência em intervalos de classes do exemplo 2.4, temos o limite inferior da terceira classe (i=3) sendo o limite inferior l_3 = 22, e o superior L_3 = 24.

Amplitude de intervalos de classe

É o tamanho do intervalo, ou seja, é a diferença entre o limite superior e o inferior, e denotamos por hi.

hi = Li – li. Na distribuição de freqüência em classes do exemplo 2.4, temos que a amplitude da terceira classe (i=3) é $h_3 = L_3 - l_3 = 24 - 22 = 2$.

Amplitude total

É a diferença entre o limite máximo e o limite mínimo do Rol. Denotamos por $AT = L_{max} - l_{min}$.

No rol do exemplo 2.3 temos que a amplitude total é AT = 28 – 17 =11 anos.

Ponto médio de uma classe

É a valor média dos extremos da classe, ou seja, é o valor que divide a classe no meio, denotamos por $P_{mi} = \frac{li + Li}{2}$, o ponto médio da classe *i*. É comum também usarmos x_i para denotarmos o ponto médio da classe *i*, mas aqui não usaremos essa notação para não confundir com os dados de uma amostra, que também denotamos por x_i. Na distribuição de frequência em classes do exemplo 2.4, temos que o ponto médio da terceira classe é $P_{m3} = \frac{l_3 + L_3}{2} = \frac{22 + 24}{2} = 23$ anos.

Frequência simples ou absoluta

É a quantidade de dados contidos em uma classe. Vamos denotar aqui por Fi, que significa a frequência absoluta, o *i* é para indicar a que classe se refere essa freqüência. Na distribuição de frequência abaixo temos que a freqüência da classe 2 é F_2=7.

i	Idade	F_i
1	17 ⊢— 19	5
2	19 ⊢— 22	7
3	22⊢—24	5
4	24 ⊢—26	2
5	26 ⊢⊣28	2
	$\sum$	21

Uma forma ainda melhor de apresentar os dados em uma distribuição de freqüência em classes, é informando as freqüências acumulativas, e percentuais. Isto facilita a análise dos dados.

Para construirmos a distribuição em intervalos de classes pelo excel utilizamos a função histograma.

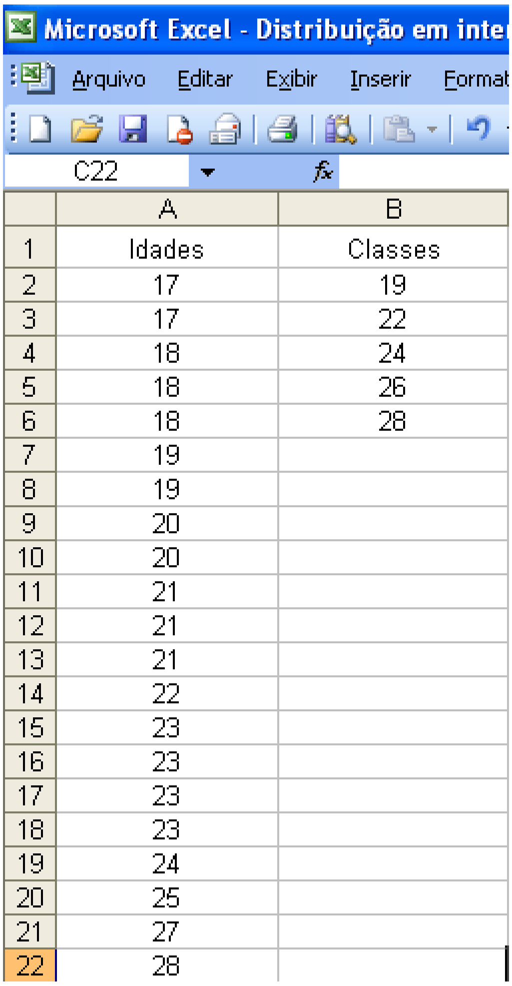
Poderíamos no exemplo acima construir nossa distribuição de freqüência em intervalos de classes tomando o primeiro intervalo fechado nos extremos e os demais fechados na direita, a planilha de cálculo excel adota essa metodologia, teríamos então:

i	Idade	F_i
1	17 ├─┤ 19	7
2	19 ─┤ 22	6
3	22 ─┤ 24	5
4	24 ─┤ 26	1
5	26 ─┤ 28	2
	$\sum$	21

Para fazermos essa distribuição em classes de freqüência utilizando o Excel 2003 devemos ir em ferramentas e escolher a opção análise de dados.

Inserimos os dados na planilha excel conforme figura abaixo, inserindo os extremos superiores dos intervalos de classes.

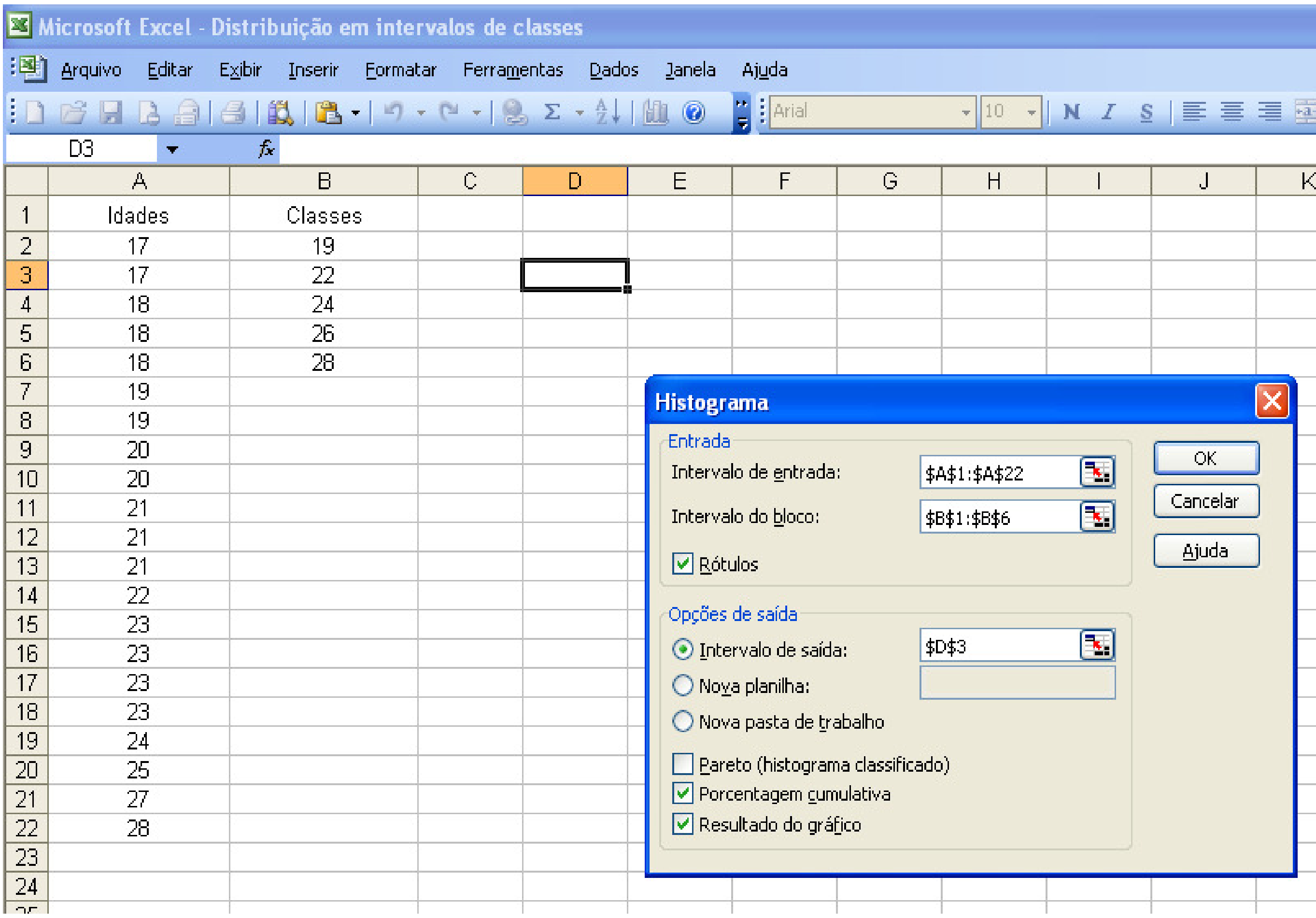
Figura 2.8



	A	B
1	Idades	Classes
2	17	19
3	17	22
4	18	24
5	18	26
6	18	28
7	19	
8	19	
9	20	
10	20	
11	21	
12	21	
13	21	
14	22	
15	23	
16	23	
17	23	
18	23	
19	24	
20	25	
21	27	
22	28	

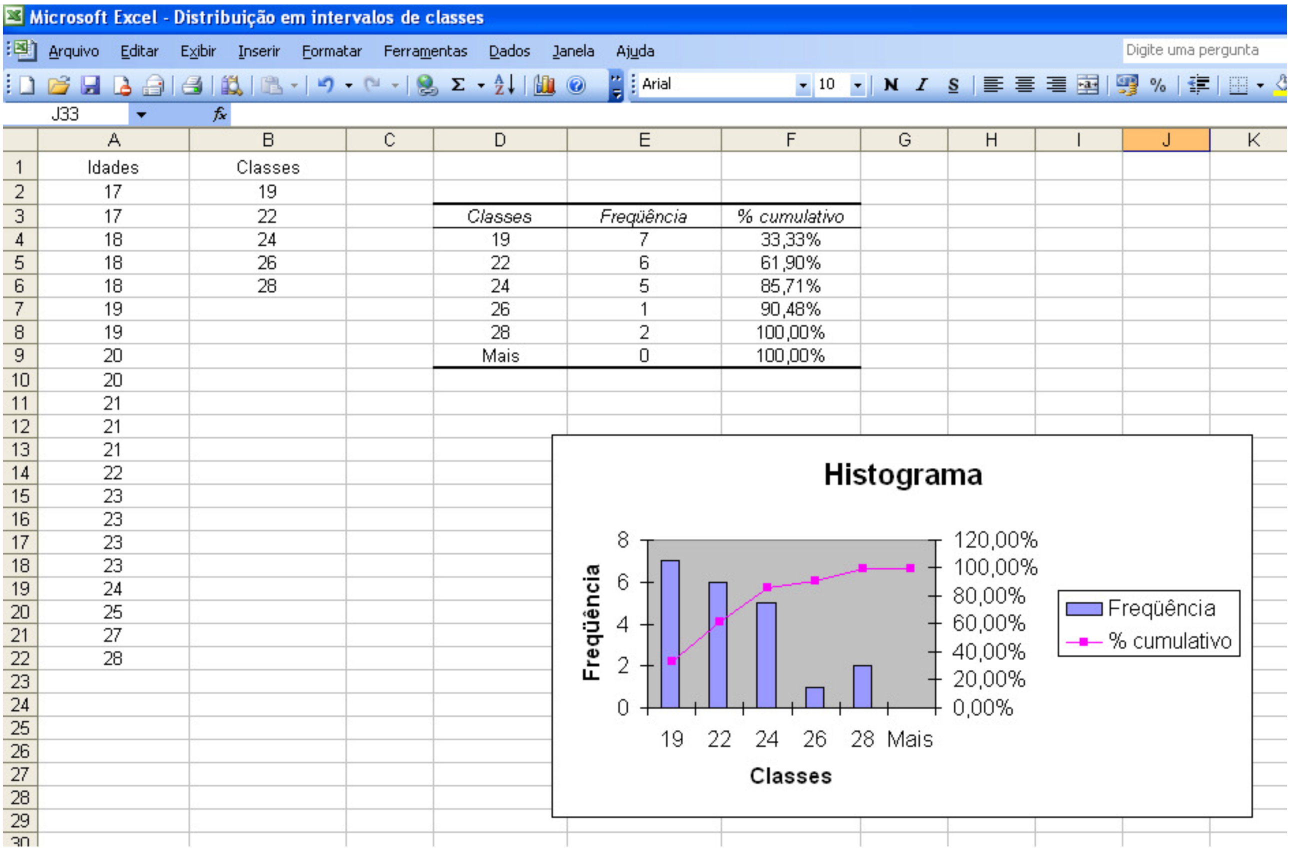
Vamos em ferramentas / análise de dados / histograma clicamos em ok e selecionamos o intervalo de entrada com o mouse de A1 a A22, o intervalo de bloco de B1 a B6, selecionamos rótulos para aparecer os títulos selecionados na distribuição em classes, escolhemos um intervalo de saída que no meu caso foi a célula D3, selecionei porcentagem cumulativa e resultado do gráfico, conforme figura 2.9:

Figura 2.9



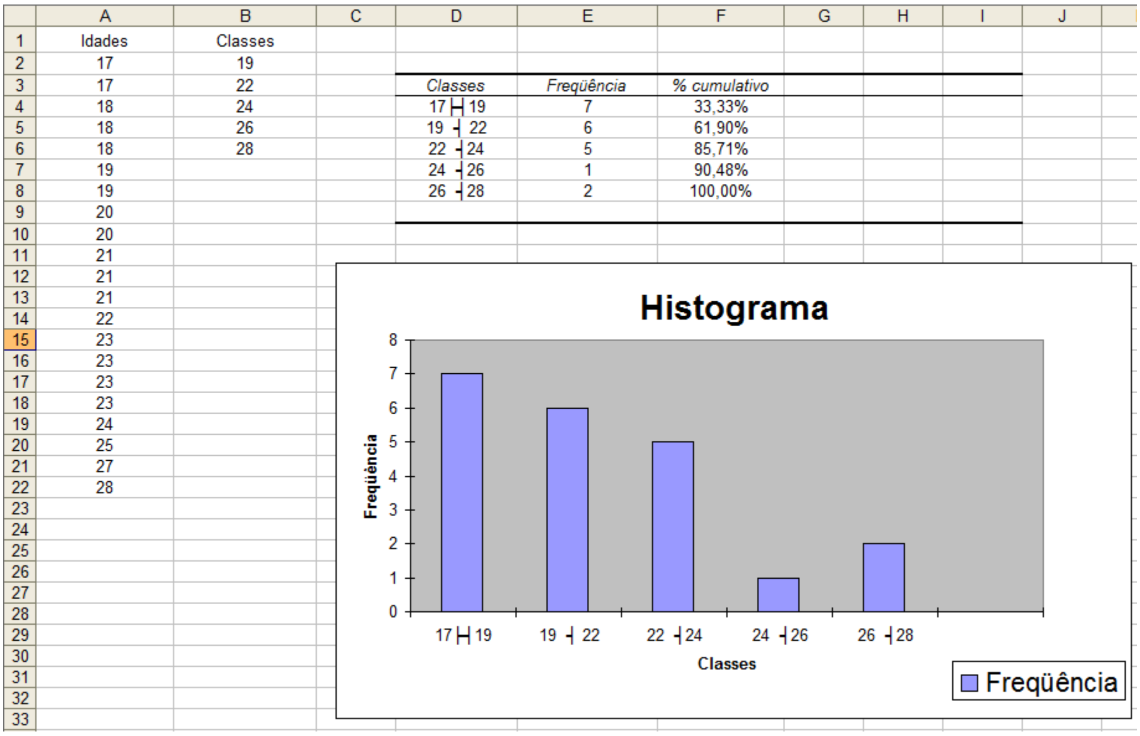
Clicamos em ok e obtemos a distribuição de freqüência e o histograma conforme figura abaixo, podemos fazer algumas arrumações na distribuição de freqüência e no gráfico utilizando o mouse, para torná-lo melhor apresentável.

Figura 2.10



Note que a distribuição em classes de freqüência que o excel construiu só aparece os limites superiores dos intervalos, mas podemos editar essa tabela para ficar melhor, inserindo os limites inferiores e os símbolos manualmente. No excel podemos inserir os símbolos com a opção inserir/símbolo. Veja figura 2.11:

Figura 2.11



No Excel 2007 o procedimento é análogo ao descrito, onde usamos a função histograma como já explicado anteriormente.

Vejamos mais um exemplo.

Exemplo 2.4: Os dados abaixo são referentes às idades de 40 alunos do ensino médio do turno noturno de uma determinada escola.

Tabela 2

i	Idade x_i	Freq F_i	Freq. Acum Fac_i	Freq. Rel $F_i\%$	Freq. Rel. Acum $Fac_i\%$
1	17	2	2	5	5
2	18	3	5	7,5	12,5
3	20	2	7	5	17,5
4	21	1	8	2,5	20
5	22	5	13	12,5	32,5
6	23	10	23	25	57,5
7	24	12	35	30	87,5
8	25	3	38	7,5	95
9	26	2	40	5	100
$\sum$		40		100	

O cálculo da freqüência relativa pode ser feito usando uma regra de três simples, ou simplesmente dividindo o número que desejamos colocar em percentual pelo total de dados e depois multiplicamos por 100%. Por exemplo, na tabela 2 acima desejamos saber quantos por cento corresponde 2 de 40, logo temos 2/40=0,05=0,05x100%=5%. As freqüências acumuladas são obtidas somando-se sempre as freqüências das classes anteriores.

2.2.5 Construção de uma distribuição de frequência em intervalos de classes utilizando a regra de Sturges

Vejamos o exemplo abaixo que evidencia o que falamos anteriormente, ou seja, quando temos dados que não se repetem com muita freqüência, a representação dos dados em uma distribuição de freqüência se torna inviável, pois teremos uma tabela muito grande o que dificulta a análise dos dados.

Exemplo 2.5: Suponha que os pesos de um grupo de estudantes sejam dados pelo seguinte Rol:

37 , 40 , 48 , 48 , 48 , 50 , 50 , 51 , 52 , 52 , 52 , 52
54 , 59 , 60 , 60 , 60 , 60 , 61 , 61 , 61 , 61 , 62 , 62
63 , 64 , 64 , 65 , 65 , 65 , 67 , 68 , 74 , 77 , 77 , 81
81 , 83 , 87 , 89

Temos o Rol acima representado na seguinte distribuição de freqüência:

Idades	F_i	$F_i\%$	Fac_i	$Fac_i\%$
37	1	2,5	1	2,5
40	1	2,5	2	5,0
48	3	7,5	5	12,5
50	2	5,0	7	17,5
51	1	2,5	8	20,0
52	4	10,0	12	30,0
54	1	2,5	13	32,5
59	1	2,5	14	35,0
60	4	10,0	18	45,0
61	4	10,0	22	55,0

62	2	5,0	24	60
63	1	2,5	25	62,5
64	2	5,0	27	67,5
65	3	7,5	30	75,0
67	1	2,5	31	77,5
68	1	2,5	32	80,0
74	1	2,5	33	82,5
77	2	5,0	35	87,5
81	2	5,0	37	92,5
83	1	2,5	38	95,0
87	1	2,5	39	97,5
89	1	2,5	40	100,0
$\sum$	40	100,0	//////////	/////////

Podemos ver que a distribuição de freqüência é grande, mesmo sendo o número de dados pequeno considerando as quantidades de dados que podemos trabalhar em uma pesquisa real. Vamos então construir a distribuição em classes de freqüência e veremos o resultado final.

Primeiro passo:

Vamos determinar o número de classes de nossa distribuição. Pela regra de Sturges, como n=40 e 40 é maior que 25 então temos que:

$$K \cong 1 + 3{,}32 \log n \Rightarrow K \cong 1 + 3{,}32 \log 40 \cong 6{,}31 \Rightarrow k = 6$$

Se utilizarmos neste caso a regra da raiz quadrada, obteremos o mesmo número de classes, ou seja, 6. Vejamos:

$$K \cong \sqrt{40} \cong 6{,}32 \Rightarrow K = 6$$

Segundo passo:

Vamos determinar a amplitude dos intervalos de classes, a amplitude dos intervalos nós denotamos por *h*.

$$h = \frac{AT}{K} = \frac{L_{\max} - l_{\min}}{K} = \frac{89 - 37}{6} = \frac{52}{6} \cong 8{,}67$$

Como os nossos dados são dados que correspondem a números inteiros, vamos então aproximar para um número inteiro, logo temos:

$$h = 9$$

Observação: Se tivéssemos, por exemplo, um conjunto de dados em que o menor dado fosse 37 e o maior fosse 87, nós teríamos:

$$h = \frac{AT}{K} = \frac{L_{\max} - l_{\min}}{K} = \frac{87 - 37}{6} = \frac{50}{6} \cong 8{,}33$$

Mas como os nossos dados são dados inteiros, nós aproximamos *h* para um número inteiro, porém devemos aproximar sempre para o inteiro mais próximo superiormente, pois na construção das classes, não teremos nenhum dado fora da última classe. Logo teríamos também *h*=9.

Se os pesos do exemplo 2.5 estivessem com três casas decimais, e tivéssemos, por exemplo, o menor dado 37,138 Kg e o maior peso 87,467 Kg, teríamos a amplitude das classes dados por:

$$h = \frac{AT}{K} = \frac{L_{\max} - l_{\min}}{K} = \frac{87{,}467 - 37{,}138}{6} = \frac{50{,}329}{6} \cong 8{,}38817$$

Mas neste caso como teríamos nossos dados com três casas decimais, teríamos de aproximar a amplitude *h* para três casas decimais, logo aproximamos *h* para *h*=*8*,389, ou seja, aproximamos para três casas decimais tomando uma aproximação elevando a última casa decimal para o mais próximo superiormente, pela mesma razão descrita acima.

Terceiro passo:

Vamos construir a distribuição de freqüência, agrupando os dados em suas respectivas classes. Mas antes vejamos o limite superior da última classe, tomando o menor valor do rol somando com o produto de *h* por *K*, ou seja, $L_k = l_{\min} + h \cdot k$, no nosso caso temos que $L_6 = 37 + 9 \cdot 6 = 91$, como o limite superior da sexta (última) classe ficou maior que maior valor do rol em 2 unidades, pois 91-89=2, é indicado que façamos a distribuição dessa diferença entre as classes, para isso, ao invés de começarmos o limite inferior da primeira classe com 37, iniciaremos com 36, assim teremos uma uniformidade melhor entre as classes.

Observação: Se tivéssemos uma diferença de um número impar e não de um número par, como foi o nosso exemplo acima, escolheríamos começar o limite inferior da primeira classe com uma unidade a mais ou a menos em relação a última classe. Por exemplo, se o nosso maior valor do rol fosse 88, teríamos uma diferença de 91-88=3, logo poderíamos começar a primeira classe com 36 ou com 35. Procedemos da mesma maneira se estivéssemos trabalhando com dados que tem casas decimais.

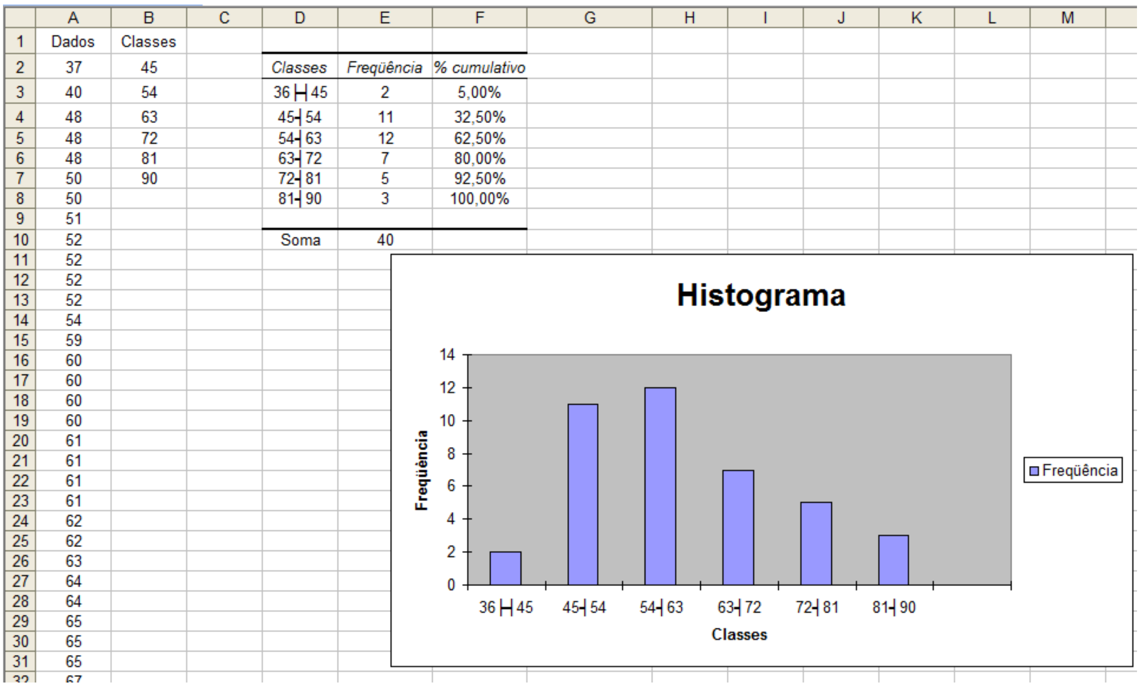
Classes	fi	fi%	Faci	FAci%
36 ⊢ 45	2	5,0	2	5,0
45 ⊢ 54	10	25,0	12	30,0
54 ⊢ 63	12	30,0	24	60,0
63 ⊢ 72	8	20,0	32	80,0
72 ⊢ 81	3	7,5	35	87,0
81 ⊢⊣ 90	5	12,5	40	100,0
∑	40	100,0	//////////////	/////////////

Ou podemos agrupar considerando os intervalos fechados à direita. Essa é a forma que a planilha de cálculo MS Excel trabalha. Vejamos:

Classes	fi	fi%	Faci	FAci%
36 ⊢⊣ 45	2	5,0	2	5,0
45 ⊣ 54	11	27,5	13	32,5
54 ⊣ 63	12	30,0	25	62,5
63 ⊣ 72	7	17,5	32	80,0
72 ⊣ 81	5	12,5	37	92,5
81 ⊣ 90	3	7,5	40	100,0
∑	40	100,0	//////////////	/////////////

Pelo excel temos:

Figura 2.12



Observação: A planilha da figura 2.12 encontra-se no arquivo distribuição em intervalos de classes, na planilha fig 2.12.

Verificamos que com os dados agrupados em classes de freqüência, facilitamos a análise dos dados.

Capítulo 3: Representação Gráfica de dados

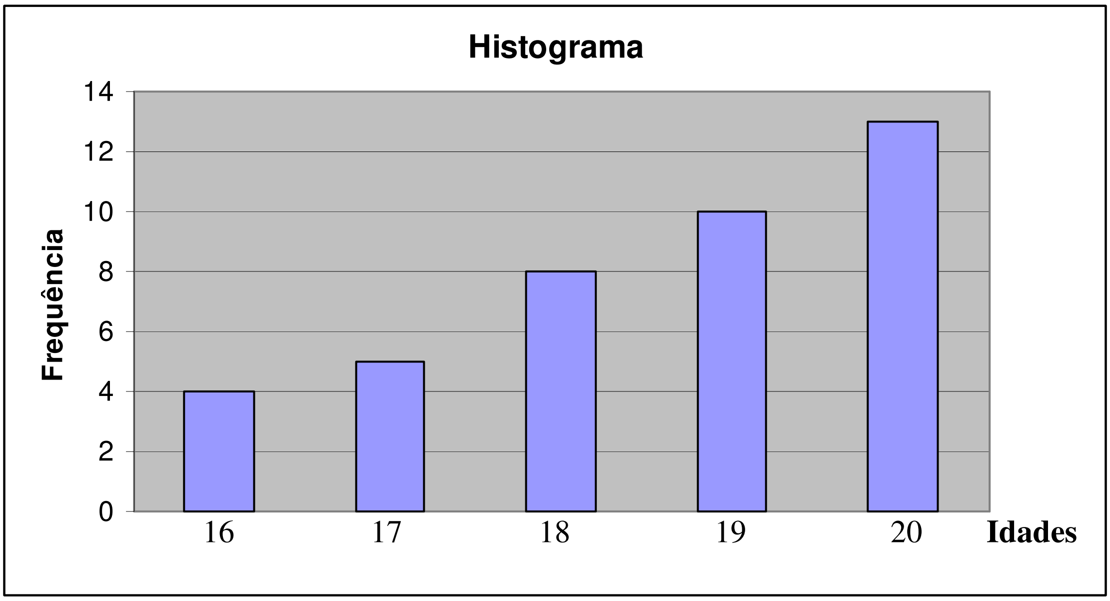
3.1 Histograma

Representa graficamente, as freqüências das classes ou os dados, que não estão agrupados em classes.

Exemplo 3.1: Vamos construir o histograma da distribuição de freqüência abaixo:

Idade	F_i	F_i%	Fac_i	Fac_i%
16	4	10	4	10
17	5	12,5	9	22,5
18	8	20,0	17	42,5
19	10	25,0	27	67,5
20	13	32,5	40	100,0
$\sum$	40	100,0	//////////	///////////

Figura 3.1

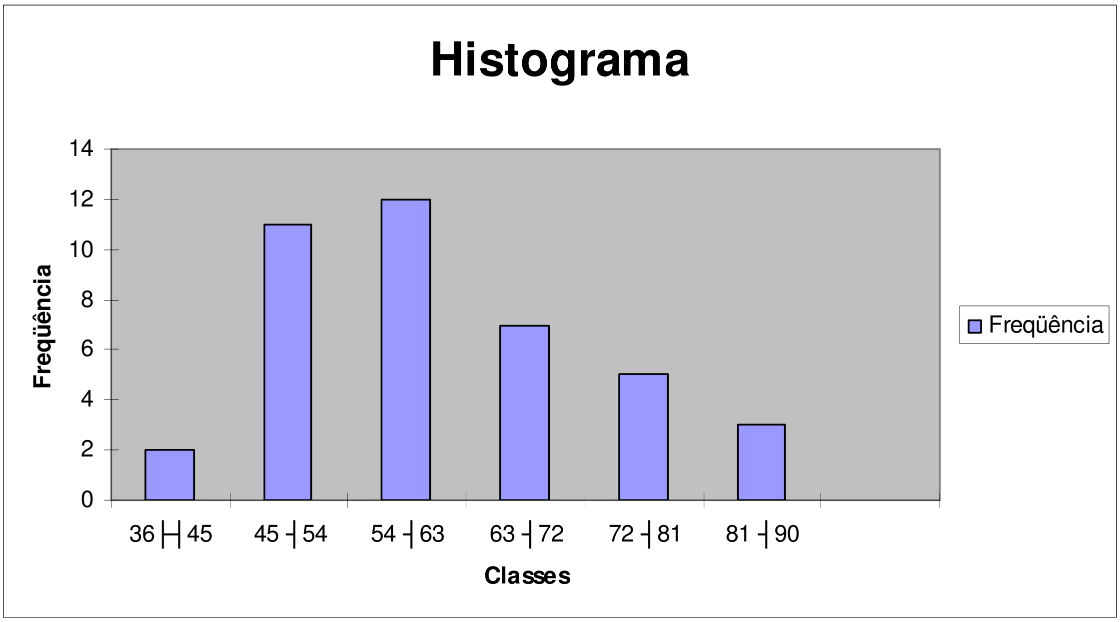


O histograma poderia ser construído evidenciando no gráfico a freqüência relativa ou até mesmo as duas juntas, isto é, a freqüência absoluta e a freqüência relativa. É comum definir histograma para distribuição de freqüência em intervalos de classes, mas aqui utilizaremos essa definição para distribuição de freqüência que não estão agrupados em classes também.

Exemplo 3.2: No caso das distribuições em intervalos de classes, o raciocínio é o mesmo, colocando no eixo horizontal as classes, como vimos no exemplo da distribuição dos pesos dos alunos que construímos anteriormente. Vejamos o histograma com evidência.

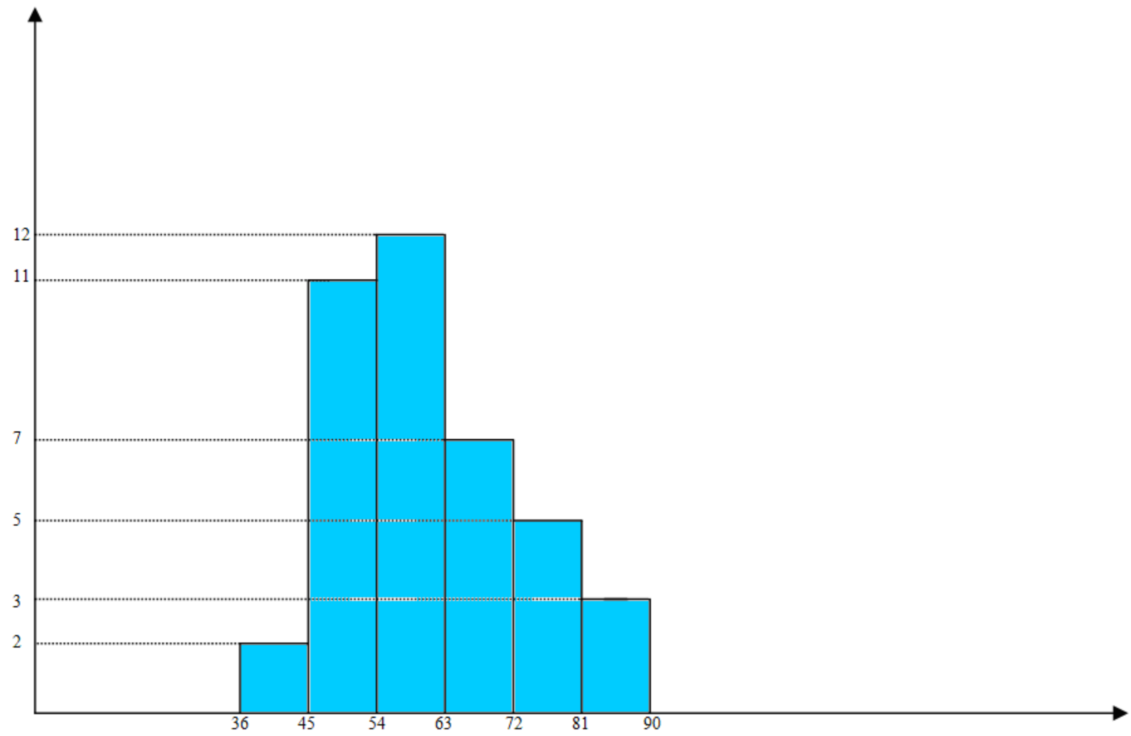
Classes	F_i	$F_i\%$	Fac_i	$Fac_i\%$
36 ├─┤ 45	2	5,0	2	5,0
45 ─┤ 54	11	27,5	13	32,5
54 ─┤ 63	12	30,0	25	62,5
63 ─┤ 72	7	17,5	32	80,0
72 ─┤ 81	5	12,5	37	92,5
81 ─┤ 90	3	7,5	40	100,0
$\sum$	40	100,0	//////////////	/////////////

Figura 3.2



O normal é construir o histograma com as colunas juntas e obedecendo uma escala para os eixos horizontal e vertical, o Excel só constrói o gráfico da forma acima, podemos então construir o gráfico em qualquer outro programa de estatística, ou mesmo em um programa de desenho, veja a figura 3.3.

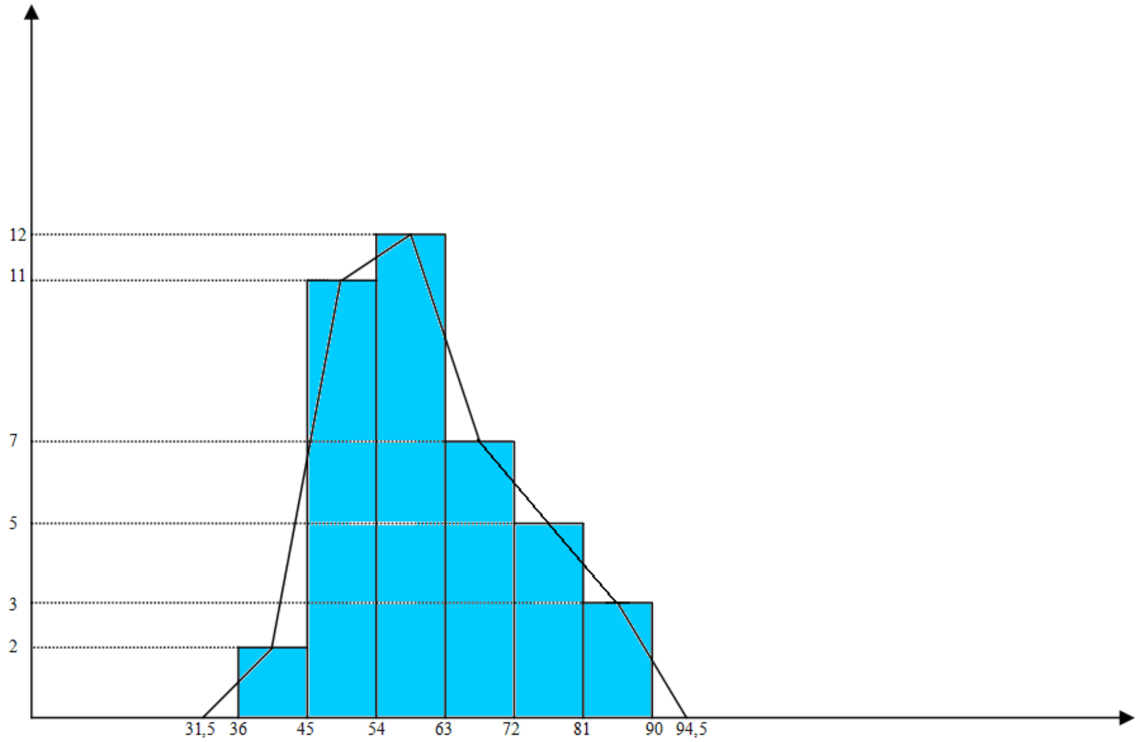
Figura 3.3



3.2 Polígono de frequência

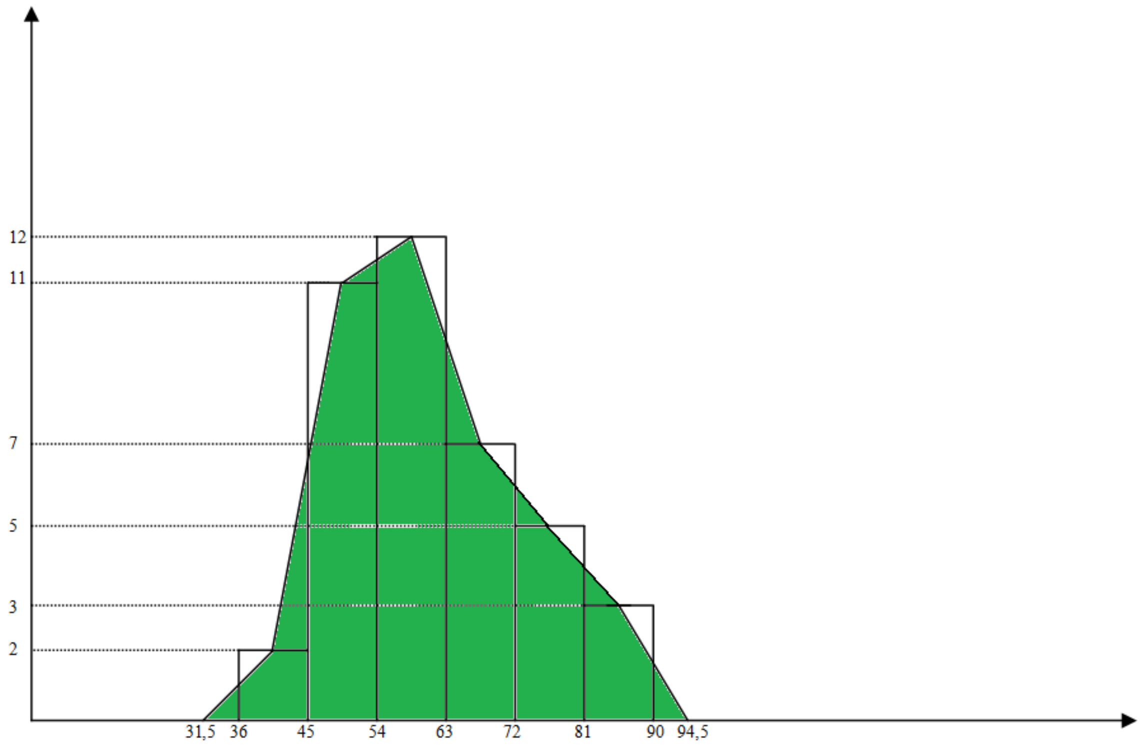
O polígono de freqüência é o polígono que obtemos quando ligamos os pontos do histograma que têm como abscissa o ponto médio de cada classe e ordenada a freqüência dessa classe, mas para construirmos o polígono de freqüência nós criamos mais dos pontos que têm ordenada zero, ou seja, está sobre o eixo horizontal, sendo o primeiro o que tem menor abscissa que é dado pela diferença do limite inferior da primeira classe com a amplitude das classes dividida por dois, isto é, $l_1 - \frac{h}{2}$, na figura 3.4 este ponto corresponde ao ponto $((l_1 - \frac{h}{2};0) = (36 - \frac{9}{2};0) = (31{,}5;0)$ e o outro ponto sendo de abscissa dada pela soma do limite superior da última classe com a metade da amplitude das classes, isto é, na figura 3.4 temos este ponto dado por $(L_6 + \frac{h}{2};0) = (90 + \frac{9}{2},0) = (94{,}5;0)$. Na figura 3.4 vemos o histograma e o polígono de freqüência.

Figura 3.4



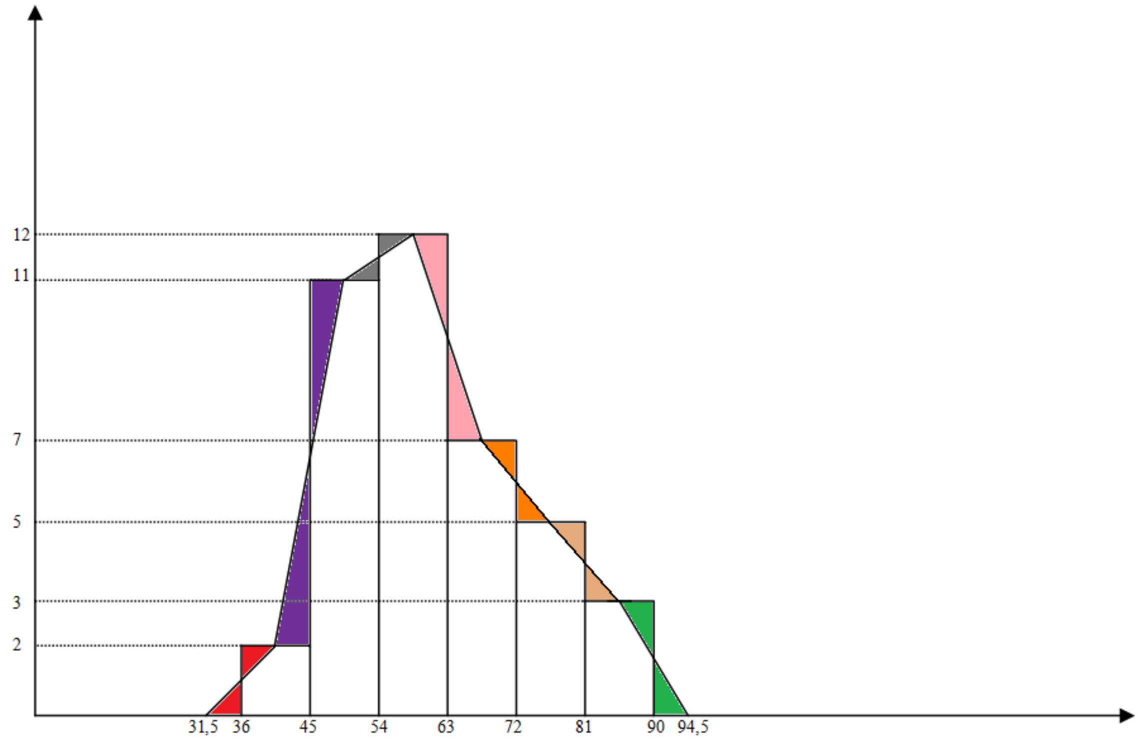
Uma observação importante é notar que a área em azul do histograma da figura 3.4, ou seja, a soma das áreas dos retângulos é igual à área limitada pelo polígono de freqüência e o eixo horizontal. Veja na figura 3.5 em verde, a área limitada pelo polígono de freqüência e o eixo horizontal.

Figura 3.5



Essas áreas são iguais, pois os triângulos de mesmas cores que estão destacados na figura 3.6 são congruentes (iguais), logo possuem as mesmas áreas. Podemos considerar a área do primeiro retângulo do histograma que representa a primeira classe tendo área $\frac{2}{40} = 0{,}05 = 5\%$, que corresponde à freqüência relativa da primeira classe, e assim sucessivamente, logo a soma das áreas dos retângulos será 1=100%. Logo a área limitada pelo polígono de freqüência e o eixo horizontal também terá área igual a 1=100%.

Figura 3.6



Observação: Para se diminuir o erro de agrupamento que surge em análises matemáticas ulteriores é aconselhável construir a distribuição em classes de freqüências, de forma que o ponto médio de cada classe coincida com dados observados.

3.3 Pareto

Uma outra representação gráfica é o pareto, que é o histograma, quando organizamos os dados ou as classes da de maior freqüência para a menor, isso facilita as comparações entre os dados ou classes de maior freqüência. No Ms Excel na ferramenta análise de dados/histograma temos na caixa de diálogo a opção pareto. Vejamos mais uma vez o exemplo dos pesos em que construímos a distribuição em classes de freqüência, e vamos construir agora no excel o pareto.

Exemplo:

Classes	fi	fi%	Faci	FAci%
36 ⊢⊣ 45	2	5,0%	2	5,0%
45 —⊣ 54	11	27,5%	13	32,5%
54 —⊣ 63	12	30,0%	25	62,5%
63 —⊣ 72	7	17,5%	32	80,0%
72 —⊣ 81	5	12,5%	37	92,5%
81 —⊣ 90	3	7,5%	40	100,0%
$\sum$	40	100,0%	//////////////	/////////////

Veja caixa de diálogo abaixo:

Figura 3.8

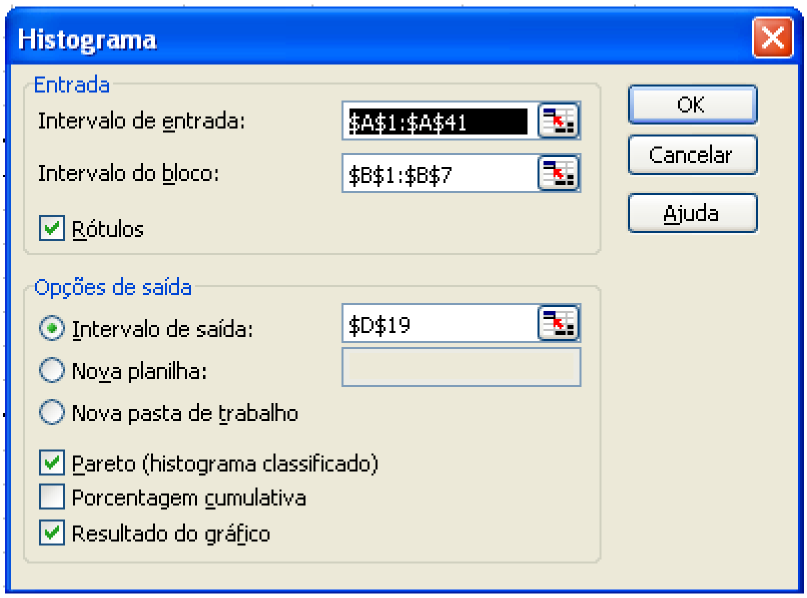
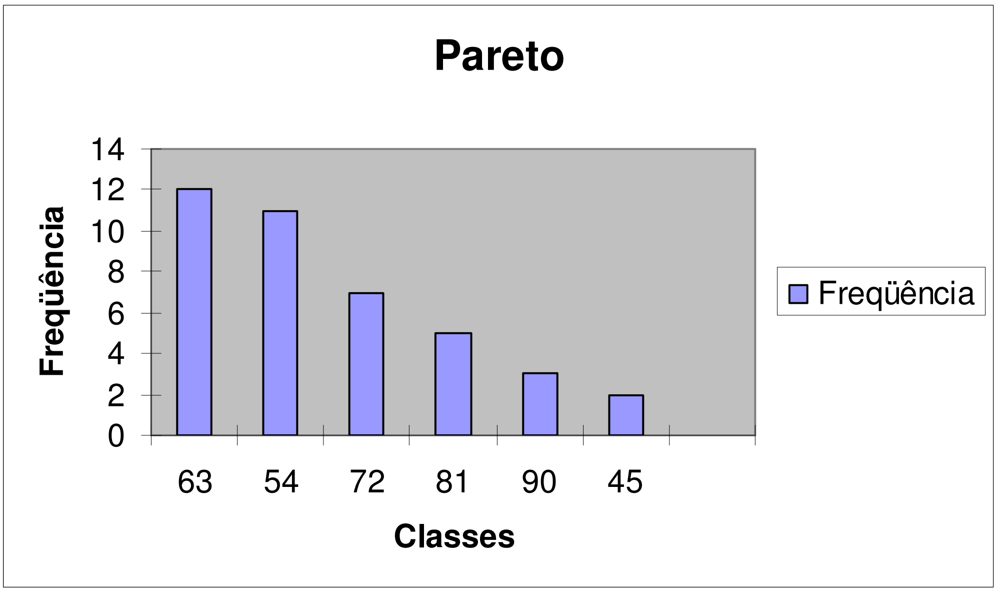


Figura 3.9



Neste caso as classes estão identificadas pelos limites superiores de cada classe, pela distribuição em classes de freqüência podemos notar que a classe de maior freqüência é a terceira classe (54 ⊢ 63) com uma freqüência de 12. No pareto essa é a classe que está em primeiro lugar e é identificada pelo limite superior dessa classe que é 63. O pareto acima encontra-se na planilha Distribuição em intervalos de classes, na planilha 3 (Plan 3).

Observação: Podemos ter uma grande quantidade de dados extraídos de uma grande população, de forma que seja possível reduzir o tamanho dos intervalos das classes, (isso é perfeitamente possível para dados contínuos, esses dados assumem qualquer valor dentro de um intervalo real), de sorte que os polígonos de freqüência tenham pequenos seguimentos de retas, de sorte que o polígono se aproxime de uma curva, que chamamos de curva de freqüência.

Vimos até aqui alguns exemplos de tratamento de dados quantitativos, isto é, de dados numéricos, vamos ver agora exemplo de tratamento de dados qualitativos, ou seja, dados que qualificam determinado assunto pesquisado.

Exemplo:

Suponha que estamos querendo investigar o interesse dos alunos pela matemática e por língua portuguesa em uma determinada escola de ensino fundamental e médio. Formulamos então um pequeno questionário com as perguntas: 1) Você gosta de matemática? 2) Você gosta de português? Onde as resposta seriam escolhidas entre as possibilidades: Não ; Pouco ; Regular ; Muito.

Questionário

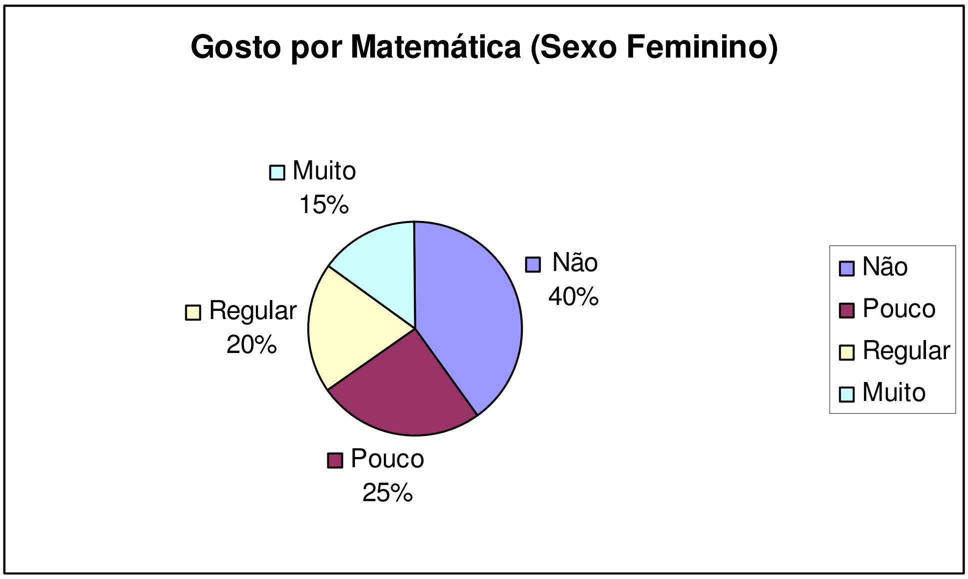
1) Você gosta de matemática?

( ) Não
( ) Pouco
( ) Regular

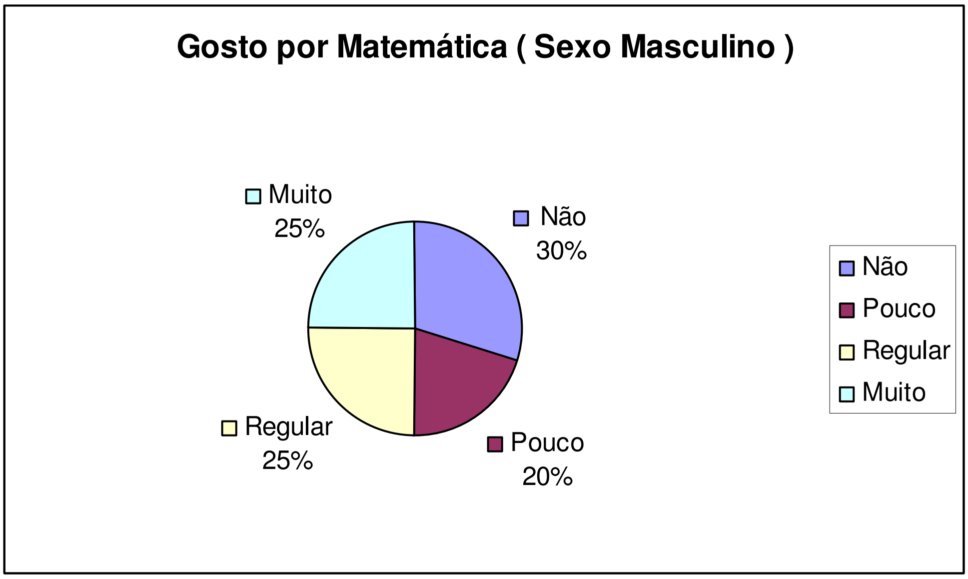
Entrevistamos então 40 alunos e construímos a seguinte distribuição de freqüência onde separamos também os alunos por sexo

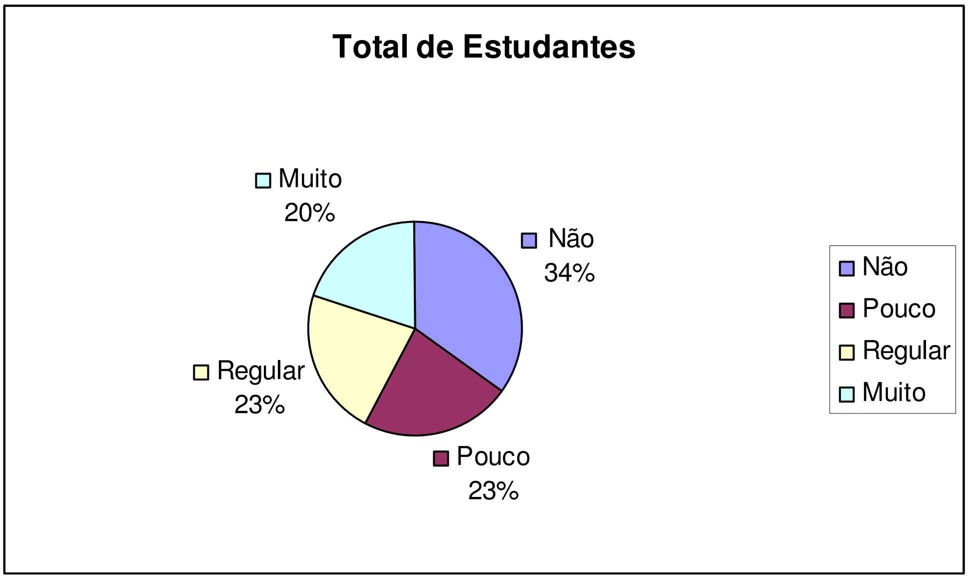
Gosto por Matemática	Feminino		Masculino		Total	
	fi	fi%	fi	fi%	fi	fi%
Não	8	40	6	30	14	35
Pouco	5	25	4	20	9	22,5
Regular	4	20	5	25	9	22,5
Muito	3	15	5	25	8	20
Total	20	100	20	100	40	100

Sugerimos como representação gráfica dessa distribuição, o gráfico de setores ou também conhecido como gráfico de pizza. Ele pode ser construído pelo MS Excel, com a função assistente de gráficos ou manualmente. Sugerimos como atividade para os alunos do ensino fundamental e médio a construção desse gráfico em papel milimetrado, onde se deve usar compasso e um transferidor para medir o ângulo dos setores, pois o ângulo dos setores devem ter as mesmas proporções que representam em relação a 360º.



Por exemplo, temos o setor do círculo em azul que representa o percentual de meninas que não gostam de matemática, logo este setor deve ter um ângulo de 40% de 360º, o que corresponde a $0,40 \cdot 360^{\circ} = 144^{\circ}$, analogamente o setor que representa pouco deve ter 90º, o setor regular 72º, e o setor muito 54º, fechando um total de 360º que corresponde ao círculo completo. Segue as representações do sexo masculino e a do total de estudantes.





CONTEÚDO 3 - Medidas de Tendência Central

Medidas de tendência central

Na estatística descritiva, uma das medidas de grande importância, são as medidads de tendência central, as medidas de tendência central resumem o comportamento central dos dados, podendo representar um conjunto de dados. Entre as medidas de tendência central, temos a média aritmética, ou simplesmente média, que é a mais famosa e mais utilizada entre as medidas de tendência central.

- Médias

    1- Média aritmética: (Média)

É calculada somando-se todos os dados e dividindo este somatório pelo número de dados.

Exemplo: Vamos calcular o peso médio (ou média dos pesos) dos 40 alunos que vimos nesse texto, esses dados encontram-se abaixo:

36 , 40 , 49 , 49 , 49 , 50 , 50 , 51 , 52 , 52 , 52 , 52
54 , 59 , 60 , 60 , 60 , 60 , 61 , 61 , 61 , 61 , 62 , 62
63 , 64 , 64 , 65 , 65 , 65 , 67 , 68 , 74 , 77 , 77 , 81
81 , 83 , 87 , 90

Somamos então os 40 pesos e dividimos esse resultado por 40

$$\text{média}=\frac{36+40+...+87+90}{40}=\frac{2474}{40}=61,85 \text{ quilos}$$

Como trabalhamos com os pesos todos com números inteiros, podemos aproximar então a média dos pesos para um número inteiro.

$$\text{Média dos Pesos} \cong 62 \text{ quilos}$$

Podemos calcular a média também pelo Ms Excel escolhendo em inserir função, a opção média. O cálculo dessa média foi feita no excel e encontra-se no arquivo medidas de Tendência Central, em planilha 1 (plan1).

Uma propriedade da média aritmética é a seguinte: Se tivermos uma série de dados em progressão aritmética (P.A), veremos que a média aritmética será igual a mediana desses dados, a mediana é um valor numérico que divide a série de dados em duas partes, uma de dados menores que a mediana e a outra em dados maiores que a mediana, logo a frente falaremos mais sobre a mediana.

Exemplo: Considere a série de dados: 5 ; 8 ; 11 ; 14 ; 17, veja que estes dados então em uma P.A de razão 3, a média desses dados é 11, e a mediana também é 11, pois esse dado divide a série no meio.

Exemplo: Considere a série de dados: 8 ; 13 ; 18 ; 23 ; 28 ; 33 veja que estes dados então em uma P.A de razão 5, a média desses dados é 20,5, e a mediana também é 20,5, pois esse valor divide a série no meio, que é exatamente um valor que está entre o dado 18 e o dado 23, tomando a média desses dois dados obtemos exatamente a mediana e a média desses dados.

Cálculo da média aritmética de dados agrupados em classes de freqüência

E se tivermos apenas a distribuição em classes de freqüência de um conjunto de dados, como faremos para calcularmos a média, se não temos todos os dados explicitados? Vamos responder essa pergunta com o exemplo abaixo:

Exemplo: Considere o Rol: 2, 4, 6, 9, 9, 10,10, 12, 13, 13, 14, 16, 17, 17, 22. Colocando estes dados em classes de freqüência temos:

Classes	f_i	Pmi
2 |- 6	2	4
6 |- 10	3	8
10 |- 14	5	12
14 |- 18	4	16
18|-|22	1	20
Σ	15	

Pois temos que n=15<25, logo o número de classes de freqüência é K=5 e a amplitude de cada classe é h=AT/K=(22-2)/5=20/5=4. Pmi é o ponto médio de cada classe. Pmi=(Li+li)/2. A média da distribuição em classes será calculada da seguinte forma:

$$\bar{x} = \frac{\sum_{i=1}^{k} f_i P_{mi}}{n},$$

Onde n é o número de dados, f_i é a freqüência da classe i, P_{mi} é o ponto médio da classe i e k é o número de classes.

logo no nosso exemplo temos que $\bar{x} = \frac{2.4+3.8+5.12+4.16+1.20}{15} = \frac{176}{15} \cong 11,73$

A média dos dados é $\bar{x} = \frac{174}{15} = 11{,}60$. Podemos ver que a média calculada pela distribuição dos dados em classes de freqüência é bem próxima da média dos dados. Logo se não tivéssemos os dados e apenas a distribuição em classes de freqüência, nós estimamos a média dos dados utilizando este cálculo para distribuição em classes de freqüência.

2- Média geométrica

Se temos n dados $x_1, x_2, \ldots, x_n$, então a média geométrica destes dados é G=$\sqrt[n]{x_1.x_2...x_n}$. A média geométrica dos pesos do exemplo anterior é calculada abaixo:

G=$\sqrt[40]{36 \cdot 40 \cdot \ldots \cdot 87 \cdot 90} \cong 60,68 \cong 61$

A média geométrica pode ser calculada pelo Ms Excel, usando a função média geométrica, veja o arquivo medidas de tendência central, em planilha 1, ou usando uma máquina científica qualquer.

Exemplo: Calcule a média geométrica dos seguintes dados 7 ; 11 ; 3 ; 5:

Temos neste caso 4 dados, então vamos calcular a raiz quarta (ou raiz de índice 4) do produto desses dados, vejamos:

G= $\sqrt[4]{7\cdot 11\cdot 3\cdot 5} \cong \sqrt[4]{1155} \cong 5{,}83$

Uma propriedade da média geométrica é a seguinte: Se tivermos uma série de dados em progressão geométrica (P.G), veremos que a média aritmética será igual a mediana desses dados

Exemplo: Considere a série de dados: 5 ; 15 ; 45 ; 135 ; 405, veja que estes dados então em uma P.G de razão 3, a média geométrica desses dados é 45, e a mediana também é 45, pois esse dado divide a série no meio.

Exemplo: Considere a série de dados: 8 ; 40 ; 200 ; 1000 ; 5000 ; 25000 veja que estes dados então em uma P.G de razão 5, a média geométrica desses dados é aproximadamente 447,21, e a mediana também pode ser 447,21, pois esse valor divide a série no meio, que é exatamente um valor que está entre o dado 200 e o dado 1000, tomando a média geométrica desses dois dados obtemos exatamente a mediana e a média geométrica desses dados.

Cálculo da média geométrica de dados agrupados em classes de freqüência

E se tivermos apenas a distribuição em classes de freqüência de um conjunto de dados, como faremos para calcularmos a média geométrica, se não temos todos os dados explicitados? Vamos responder essa pergunta com o exemplo abaixo:

Exemplo: Considere o Rol: 2, 4, 6, 9, 9, 10,10, 12, 13, 13, 14, 16, 17, 17, 22. Colocando estes dados em classes de freqüência temos:

Classes	f_i	Pmi
2 |- 6	2	4
6 |- 10	3	8
10 |- 14	5	12
14 |- 18	4	16
18|-|22	1	20
Σ	15	

A média geométrica da distribuição em classes será calculada da seguinte forma:

$$G=\sqrt[n]{P_{m1}^{\ f_1}\cdot P_{m2}^{\ f_2}\cdots P_{mk}^{\ f_k}}\ ;$$

Onde *n* é o número de dados, f_i é a freqüência da classe *i*, P_{mi} é o ponto médio da classe *i* e *k* é o número de classes.

Logo no nosso exemplo temos que:

$$G=\sqrt[15]{4^2\cdot 8^3\cdot 12^5\cdot 16^4\cdot 20^1}\cong 10,68$$

A média geométrica dos dados é

$$G=\sqrt[15]{2\cdot 4\cdot 6\cdot 9\cdot 9\cdot 10\cdot 10\cdot 12\cdot 13\cdot 13\cdot 14\cdot 16\cdot 17\cdot 17\cdot 22}=$$
$$=\sqrt[15]{2\cdot 4\cdot 6\cdot 9^2\cdot 10^2\cdot 12\cdot 13^2\cdot 14\cdot 16\cdot 17^2\cdot 22}\cong 10,08$$

Podemos ver que a média geométrica calculada pela distribuição dos dados em classes de freqüência é bem próxima da média geométrica dos dados. Logo se não tivéssemos os dados e apenas a distribuição em classes de freqüência, nós estimamos a média geométrica dos dados utilizando este cálculo para distribuição em classes de freqüência.

3- Média harmônica

Se temos n dados $x_1, x_2, \ldots, x_n$, a média harmônica destes dados é calculada da seguinte maneira: $\bar{x}_h=\dfrac{n}{\dfrac{1}{x_1}+\dfrac{1}{x_2}+\ldots+\dfrac{1}{x_n}}=\dfrac{n}{\sum\dfrac{1}{x_i}}$.

Exemplo: Vamos calcular a média harmônica dos pesos dos estudantes

$$\bar{x}_h = \frac{40}{\frac{1}{36}+\frac{1}{40}+...+\frac{1}{87}+\frac{1}{90}} \cong 59,50 \cong 60$$

Uma aplicação importante da média harmônica é no cálculo de velocidade média quando as distâncias percorridas são iguais.

Exemplo: Suponha que um motorista foi de Salvador para Feira de Santana à velocidade média de 80Km/h, e voltou de Feira de Santana para Salvador pelo mesmo caminho à velocidade média de 100Km/h. Sabendo-se que a distância entre as cidades é de 100Km. Qual foi a velocidade média de todo percurso?
O tempo que o motorista levou para ir de Ssa para Feira foi $\frac{100km}{80Km/h} = 1,25h$. O tempo que ele levou para voltar foi $\frac{100Km}{100Km/h} = 1h$. Como o percurso completo tem 200Km a velocidade média foi de $\frac{200Km}{1h+1,25h} = \frac{200Km}{2,25h} = 88,89Km/h$. Se nós calculássemos a média das velocidades médias o resultado estaria errado, pois $\frac{80Km/h+100Km/h}{2} = 90Km/h$. Notamos que neste caso temos a média harmônica das velocidades médias, pois $\frac{2}{\frac{1}{80Km/h}+\frac{1}{100Km/h}} = \frac{2}{\frac{5+4}{400Km/h}} = \frac{2}{\frac{9}{400Km/h}} = \frac{800Km/h}{9} = 88,89Km/h$.

Cálculo da média harmônica de dados agrupados em classes de freqüência

E se tivermos apenas a distribuição em classes de freqüência de um conjunto de dados, como faremos para calcularmos a média harmônica, se não temos todos os dados explicitados? Vamos responder essa pergunta com o exemplo abaixo:

Exemplo: Considere o Rol: 2, 4, 6, 9, 9, 10,10, 12, 13, 13, 14, 16, 17, 17, 22. Colocando estes dados em classes de freqüência temos:

Classes	f_i	Pmi
2 |- 6	2	4
6 |- 10	3	8
10 |- 14	5	12
14 |- 18	4	16
18|-|22	1	20
Σ	15	

A média harmônica da distribuição em classes será calculada da seguinte forma:

$$\bar{x}_h = \frac{n}{f_1 \cdot \frac{1}{P_{m1}} + f_2 \cdot \frac{1}{P_{m2}} + \ldots + f_k \cdot \frac{1}{P_{mk}}};$$

Onde n é o número de dados, f_i é a freqüência da classe i, P_{mi} é o ponto médio da classe i e k é o número de classes.

Logo no nosso exemplo temos que:

$$\bar{x}_h = \frac{15}{2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{8} + 5 \cdot \frac{1}{12} + 4 \cdot \frac{1}{16} + 1 \cdot \frac{1}{20}} \cong 9,42$$

A média harmônica dos dados é

$$\bar{x}_h = \frac{15}{\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + \frac{1}{9} + \frac{1}{9} + \frac{1}{10} + \frac{1}{10} + \frac{1}{12} + \frac{1}{13} + \frac{1}{13} + \frac{1}{14} + \frac{1}{16} + \frac{1}{17} + \frac{1}{17} + \frac{1}{22}} =$$

$$= \frac{15}{\frac{1}{2} + \frac{1}{4} + \frac{1}{6} + 2 \cdot \frac{1}{9} + 2 \cdot \frac{1}{10} + \frac{1}{12} + 2 \cdot \frac{1}{13} + \frac{1}{14} + \frac{1}{16} + 2 \cdot \frac{1}{17} + \frac{1}{22}} \cong 9,03$$

Podemos ver que a média harmônica calculada pela distribuição dos dados em classes de freqüência é bem próxima da média harmônica dos dados. Logo se não tivéssemos os dados e apenas a distribuição em classes de freqüência, nós estimamos a média harmônica dos dados utilizando este cálculo para distribuição em classes de freqüência.

- Mediana

É o valor que divide o rol no meio, isto é, uma parte dos dados, contém dados menores que a mediana, e outra parte com a mesma quantidade de dados, contém dados maiores que a mediana.

Exemplo: Doze candidatos fizeram uma prova seletiva para o preenchimento de seis vagas para administrador em uma empresa. As notas foram as seguinte: 5,0 ; 7,0 ; 6,5 ; 6,0 ; 8,0 ; 7,0 ; 5,5 ; 9,0 ; 9,5 ; 8,5 ; 8,0 ; 7,6. Sabendo-se que a nota de corte é a mediana. Determine as notas dos candidatos selecionados.

Devemos colocar os dados em ordem crescente, que é exatamente o rol dos dados

Rol: 5,0 ; 5,5 ; 6,0 ; 6,5 ; 7,0 ; 7,0 ; 7,6 ; 8,0 ; 8,0 ; 8,5 ; 9,0 ; 9,5

Como temos um número par de dados, tomamos a mediana sendo a média dos dados centrais, que neste caso são x_6=7,0 e x_7=7,6, logo M_d=(7+7,6)/2=7,3. Para determinarmos os dados centrais de uma série como número de dados par, nós dividimos o total de dados por dois, para acharmos o primeiro dado central, o outro dado central é o seguinte. No exemplo acima fizemos 12/2=6, então x_6 é o primeiro dado central e x_7 é o segundo dado central, logo $M_d=(x_6+x_7)/2$.

$$M_d=\frac{7,0+7,6}{2}=\frac{14,6}{2}=7,3$$

Exemplo: Suponha que treze candidatos fizeram uma prova seletiva para o preenchimento de seis vagas para administrador em uma empresa. As notas foram as seguintes: 5,0 ; 7,0 ; 6,5 ; 6,0 ; 8,0 ; 7,0 ; 5,5 ; 9,0 ; 9,5 ; 8,5 ; 8,0 ; 7,6 ; 8,7. Sabendo-se que a nota de corte é a mediana. Determine as notas dos candidatos selecionados.

Rol: 5,0 ; 5,5 ; 6,0 ; 6,5 ; 7,0 ; 7,0 ; 7,6 ; 8,0 ; 8,0 ; 8,5 ; 8,7 ; 9,0 ; 9,5

Como a quantidade de dados é um número impar, temos então que a mediana vai ser um dado do rol, para localizarmos a posição deste dado no rol, nós somamos a quantidade de dados com um e dividimos por dois. No exemplo anterior tínhamos um total de 13 dados, logo (13+1)/2 = 14/2=7, logo a posição do dado central no rol é a sétima posição, ou seja, é o x_7 = 7,6, que é justamente a mediana.

$$M_d = 7{,}6$$

Podemos calcular a mediana pelo Ms Excel, utilizando a função Med (mediana) na categoria estatística, veja o arquivo Medidas de tendência Central, em planilha 2 (plan 2)

Cálculo da mediana de dados agrupados em classes de freqüência

E se tivermos apenas a distribuição em classes de freqüência de um conjunto de dados, como faremos para calcularmos a mediana, se não temos todos os dados explicitados? Vamos responder essa pergunta com o exemplo abaixo:

Exemplo: Considere o Rol: 2, 4, 6, 9, 9, 10,10, 12, 13, 13, 14, 16, 17, 17, 22. Colocando estes dados em classes de freqüência temos:

Classes	f_i	$f_i\%$	Fac_i	$Fac_i\%$
2 |- 6	2	13,33%	2	13,33%
6 |- 10	3	20,00%	5	33,33%
10 |- 14	5	33,33%	10	66,66%
14 |- 18	4	26,67%	14	93,33%
18|-|22	1	6,67%	15	100,00%
Σ	15	100,00%		

A idéia para se calcular a mediana de uma distribuição em classes é em primeiro lugar localizar em que classe se encontra a mediana, na distribuição acima temos 15 dados, sabemos que a mediana dos dados é o dado que está na oitava posição do rol, logo vemos que a classe onde se encontra esse oitavo dado é a terceira classe, (10 |- 14), se por acaso tivéssemos uma quantidade par de dados, e se um dos dados centrais caísse em uma classe e o outro na classe seguinte, poderíamos considerar uma das duas como classe da mediana. Sabemos então que neste exemplo a classe da mediana é a terceira classe, logo a mediana tem que ser um valor que está neste intervalo, para isso vamos tomar o limite inferior dessa classe e somar a um percentual da amplitude desse intervalo, esse percentual é calculado tomando a metade da quantidade de dados que depois subtraímos da freqüência acumulada da classe anterior a classe da mediana e depois dividimos pela freqüência da classe da mediana, para acharmos esse percentual relativo a quantidade de dados da classe da mediana. Podemos escrever isso em uma expressão.

$$M_d = l + h\frac{(E_{Md} - F_{ant})}{f_{Md}}$$; onde

l: É o limite inferior da classe da mediana.
h: è a amplitude da classe de mediana.
E_{Md}: É a freqüência total dividida por dois.

F_{ant}: É a freqüência acumulada da classe anterior à classe da mediana.
f_{Md}: É a freqüência da classe da mediana.

Logo no nosso exemplo temos que l=10, h=4, E_{Md}=7,5, F_{ant}=5 e f_{Md}=5, desta forma temos que

$$M_d = 10 + 4 \cdot \frac{(7,5-5)}{5} = 12 .$$

Podemos usar nos cálculos as freqüências relativas que vamos obter a mesma resposta, vejamos:

$$M_d = 10 + 4 \cdot \frac{(50\% - 33,33\%)}{33,33\%} = 12$$

Neste nosso exemplo coincidentemente obtemos que a mediana da distribuição em classe é a mesma mediana dos dados, pois o oitavo dado do rol é exatamente 12.

Vamos ver mais um exemplo:

Considere a distribuição em classes abaixo

Classes	fi	fi%	Faci	Faci%
2 |- 4,4	3	30	3	30
4,4 |- 6,8	1	10	4	40
6,8 |- 9,2	2	20	6	60
9,2 |- 11,6	2	20	8	80
11,6 |-14	2	20	10	100
$\sum$	10	100	///////	/////////

Temos neste caso que os dados centrais então na terceira classe, que vai de 6,8 e 9,2. Os dados centrais são o x_5 e x_6, pois temos dez dados. Neste caso temos que, l=6,8, h=2,4, E_{Md}=50, F_{ant}=40 e f_{Md}=20, desta forma temos que $M_d=6,8+2,4.\frac{(50-40)}{20}=8$. Se usarmos os dados numéricos ao invés de porcentagens temos que, E_{Md}=5, F_{ant}=4 e f_{Md}=2, desta forma temos que $M_d=6,8+2,4.\frac{(5-4)}{2}=8$

- Moda

A moda de um conjunto de dados é o dado que mais se repetiu, ou seja, é o dado com a maior freqüência, logo a moda de um conjunto de dados pode não existir e neste caso a série de dados é dita amodal. Podemos ter também uma série de dados com mais de uma moda, dizemos neste caso que a série é multimodal. A moda é a única medida de tendência central que se aplica em dados quantitativos e qualitativos.

Exemplo (*quantitativo*): O número de infrações de trânsito observadas em três semáforos A, B e C de uma cidade, a cada hora, durante 8 horas de observação foram os seguintes:

A: 10, 12, 14, 14, 16, 8, 9, 5
B: 15, 20, 20, 12, 13, 15, 10, 2
C: 5, 6, 4, 2, 3, 1, 7, 8

Moda de A=14
Moda de B = 20 e 15
Moda de C não existe.

Exemplo (*qualitativo*): Foi feita uma pesquisa com 8 pessoas, que assistiram três filmes, A, B e C, onde elas tinham que classificar como P (péssimo), R (regular), B (bom) ou O (ótimo) e obtemos os seguintes resultados:

Filme A: B, B, R, R, O, O, O, O — Moda filme A = O
Filme B: P, P, R, R, R, B, B, B — Moda filme B = R e B
Filme C: P, P, R, R, B, B, O, O — Moda filme C não existe

Exemplo: Uma fábrica de calças fez uma pesquisa com mil pessoas do sexo masculino de uma cidade, para saber o número mais comum que estas pessoas vestiam. De acordo com a tabela de freqüência abaixo, determine a moda dos dados.

Número da calça	f_i
36	200
38	250
40	350
42	500
44	300
$\sum$	1600

A moda dos dados é 42, que é exatamente o número mais usado na cidade.

Podemos calcular a moda de um conjunto de dados utilizando o Ms Excel, pela função MODO da categoria estatística. Para dados qualitativos podemos usar a função do Ms Excel CONT.SE. Veja arquivo medidas de tendência central, em planilha 2.

CONTEÚDO 4 - Medidas de Dispersão

Medidas de dispersão

O grau ao qual os dados numéricos tendem a dispersar-se em torno de um valor médio, nós chamamos de dispersão ou variação dos dados.

- Amplitude total (h)

É a diferença entre o maior valor e o menor valor de um conjunto de dados.

Exemplo: Considere o conjunto A={5, 2, 3, 9, 1, 7}. O Rol deste conjunto é: 1, 2, 3, 5, 7, 9. A Amplitude total destes é h=9-1=8

- Desvio médio (DM)

É o somatório dos módulos da diferença entre cada elemento do conjunto e a média deste conjunto dividido pelo número de dados.

Exemplo: Considere o conjunto de dados {2, 4, 9}. Temos que a média destes dados é $\bar{x} = 5$. Logo DM=$\frac{|2-5|+|4-5|+|9-5|}{3} = \frac{|-3|+|-1|+|4|}{3} = \frac{3+1+4}{3} = \frac{8}{3} = 2{,}67$. Se temos um conjunto com n dados {x_1, x_2, ... , x_n}, com média $\bar{x}$, então DM=$\frac{\sum |x_i - \bar{x}|}{n}$.
O desvio médio pode ser calculado pela planilha MS. Excel utilizando a função Desvio Médio (DESV.MÉDIO).

- Variância Populacional (σ^2)

É o somatório do quadrado da diferença entre cada elemento da população, e a média destes elementos, dividido pelo número de elementos.

Exemplo: Considere o conjunto de dados {1, 3, 2, 6}, temos que a média dos dados é $\bar{x} = 3$ e a variância

$$\sigma^2 = \frac{(1-3)^2 + (3-3)^2 + (2-3)^2 + (6-3)^2}{4} = \frac{(-2)^2 + 0^2 + (-1)^2 + 3^2}{4} = \frac{4+0+1+9}{4} = \frac{14}{4} = 3{,}50$$
.

Se temos um conjunto com n dados {x_1, x_2, ... , x_n}, então σ^2 = $\frac{\sum (x_i - \bar{x})^2}{n}$.

A variância populacional pode ser calculado pela planilha MS. Excel utilizando a função variância populaçional (VARP).

- Desvio padrão populacional (σ)

É a raiz quadrada da variância populacional, σ = $\sqrt{\sigma^2} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$.

Exemplo: Considere o exemplo anterior. Obtemos a variância dos dados σ^2 = 3,50, logo o desvio padrão dos dados é $\sigma = \sqrt{\sigma^2} = \sqrt{3{,}50} = 1{,}87$.

O desvio padrão populacional pode ser calculado pela planilha MS. Excel utilizando a função desvio padrão populaçional (DESVPADP).

- Desvio padrão e variância populacional e amostral

O desvio padrão e a variância podem ser calculados de forma amostral, isto é, tomando-se uma parte da população, e neste caso, são ditos amostrais, ou podem ser calculados usando-se toda população, e neste caso são ditos populacionais.

Variância populacional: $\sigma^2 = \dfrac{\sum (x_i - \bar{x})^2}{n}$

Desvio padrão populacional: $\sigma = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n}}$

Variância amostral: $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1}$

Desvio padrão amostral: $s = \sqrt{\dfrac{\sum (x_i - \bar{x})^2}{n-1}}$

Exemplo: A variância populacional dos dados do exemplo anterior foi σ^2=3,50, e o desvio populacional foi σ=1,87. Se esses quatro dados refere-se a uma amostra de uma população, então temos a variância amostral sendo: $s^2 = \dfrac{\sum (x_i - \bar{x})^2}{n-1} \; \dfrac{(1-3)^2 + (3-3)^2 + (2-3)^2 + (6-3)^2}{3} = \dfrac{(-2)^2 + 0^2 + (-1)^2 + 3^2}{3} == \dfrac{14}{3} = 4{,}67$

e o desvio padrão amostra é $s = \sqrt{s^2} = \sqrt{4{,}67} = 2{,}16$. Das medidas de dispersão as mais usadas são o desvio padrão amostral e a variância amostral.

Exercício: Calcule a média, o desvio padrão e a variância dos seguintes dados: 5,0 ; 5,5 ; 6,0 ; 6,5 ; 7,0 ; 7,0 ; 7,6 ; 8,0 ; 8,0 ; 8,5 ; 9,0 ; 9,5.

Resposta: $\bar{x} = 7{,}3$, s=1,39 e s^2=1,93.

Coeficiente de dispersão relativa (dispersão relativa)

Obtemos a dispersão relativa dividindo a dispersão absoluta pela média, onde a dispersão absoluta pode ser qualquer medida de dispersão, logo *dispersão relativa* = $\dfrac{\text{dispersão absoluta}}{\text{média}}$.

A razão entre o desvio padrão e a média de uma série é dito coeficiente de dispersão relativa, logo *coeficiente de dispersão relativa* = $\frac{\sigma}{\bar{x}}$ ou *coeficiente de dispersão relativa* = $\frac{s}{\bar{x}}$.

Exemplo: A média de uma turma na prova de estatística foi 9 e em matemática foi 8, sabendo-se que a turma tem 30 alunos e que o desvio padrão das notas de estatística foi σ =4 e em matemática foi σ =5. Determine o coeficiente de dispersão relativa de cada matéria.

Estatística: *Coeficiente de dispersão relativa* = $\frac{\sigma}{\bar{x}} = \frac{4}{9} = 0{,}44 = 44\%$

Matemática: : *Coeficiente de dispersão relativa* = $\frac{\sigma}{\bar{x}} = \frac{5}{8} = 0{,}63 = 63\%$

A matéria matemática apresentou um coeficiente de dispersão relativa maior.

Medidas de ordenamento e posição

- Quartis: Sabemos que a mediana divide o Rol em duas partes, sendo 50% dos dados menores que a mediana e 50% dos dados maiores que a mediana. O quartis divide o Rol em quatro partes iguais. Logo teremos três quartis Q_1, Q_2 e Q_3, que dividiram o Rol em quatro partes iguais. Para determinarmos a ordem dos quartis, nós usaremos a relação Q_i = $x_{[\frac{i.n}{4}+\frac{1}{2}]}$, onde i = 1,2 e 3 são as ordens dos quartis e n é o número de dados.

Exemplo: Considere os dados 11 ; 7 ; 5 ; 19 ; 17 ; 13. Determine os quartis deste conjunto de dados.

Rol: 5 ; 7 ; 11 ; 13 ; 17 ; 19. Neste caso temos seis dados, logo n=6. O primeiro quartil é $Q_1 = x_{[\frac{1.6}{4}+\frac{1}{2}]} = x_2 = 7$. O segundo e o terceiro quartis são $Q_2 = x_{[\frac{2.6}{4}+\frac{1}{2}]} = x_{3,5} = \frac{x_3 + x_4}{2} = \frac{24}{2} = 12$ e $Q_3 = x_{[\frac{3.6}{4}+\frac{1}{2}]} = x_5 = 17$. Logo estes são os três quartis que dividem o Rol em quatro partes iguais.

$$\begin{array}{cccccc} 5 & 7 & 11 & 13 & 17 & 19 \\ & \uparrow & \uparrow & \uparrow & & \\ & Q_1 & Q_2 & Q_3 & & \end{array}$$

- Decis: Dividem o Rol em dez partes iguais. $D_i = x_{[\frac{i.n}{10}+\frac{1}{2}]}$.
- Percentis: Dividem o Rol em cem partes iguais. $P_i = x_{[\frac{i.n}{100}+\frac{1}{2}]}$.
-

Exercício: Dado o conjunto de dados {1,3,2,6,5,9}. Determinar seus quartis.

Resposta: $Q_1=2$, $Q_2=4$, $Q_3=6$

Exercício: Dado o conjunto de dados {20, 5, 7, 3, 9, 3, 4, 8, 2}, determine:

a) O Rol, a média e a mediana.
b) A moda, o desvio médio e a variância.
c) O desvio padrão , o quartil Q_3, o decil D_5 e o percentil P_{10}

Respostas: a) $\bar{x} = 6{,}78$, $M_d=5$
b) Moda = 3, DM=3,75 e s^2 =30,44
c) s = 5,52, $Q_3=x_{7,75}=8{,}75$, $D_5=x_5=7$, $P_{10}=x_{1,4}=2{,}4$.

Exercício: Considere os conjuntos de dados D1={13,29,37,51,46,39,58} e D2={14,26,13,32,16,53,78,86,93,41}. Determine para os dois conjuntos de dados:

a) A mediana, a média, e a moda
b) Q_3, D_7 e P_{52}

Atividade Complementar

1. Dado o rol de dados amostrais abaixo:

26,98	25,10	25,20	32,40	34,30	34,30	36,40	39,06	39,37
39,40	49,00	49,00	50,84	51,50	54,68	54,68	65,30	65,60
67,70	67,80	70,30	71,77	71,80	79,60	79,70	81,00	81,00
81,00	82,37	82,38	89,00	93,00	93,90	94,13	101,00	103,00
107,00	107,00	107,00	107,30	120,34	122,87	122,98	123,96	

a) Construa a distribuição em classes de freqüência.
b) Construa o histograma em papel milimetrado e no Ms. Excel
c) Calcule a média dos dados com o auxilio de uma máquina de calcular e no Ms. Excel
d) Determine a mediana dos dados manualmente e com o Ms. Excel.

2. Em cada uma das situações abaixo utilize o cálculo da média correta. Escolha entre o cálculo da média aritmética, da média geométrica ou da média harmônica. Justifique a utilização de cada média em cada situação.

a) Um motorista foi de Lauro de Freitas para Camaçarí à velocidade média de 78,92 km/h. Depois foi de Camaçarí para Feira de Santana à velocidade média de 103,78 km/h. Depois voltou de Feira de Santana para Salvador à velocidade média de 69,23 km/h. Calcule a velocidade média de todo este percurso, supondo que as distância entre as cidades são iguais.

b) A média das notas da turma dos estudantes de administração com habilitação em marketing foi 7,98 na prova de estatística, a média das notas dos estudantes de administração com habilitação em comercio exterior foi de 7,98 e a dos estudantes de administração com habilitação em gestão da informação foi de 8,38. Sabendo-se que temos 41, 43 e 36 alunos nas turmas de marketing, comercio exterior e gestão da informação respectivamente, calcule a média das notas de todos os alunos das três habilitações.

3. As notas de 17 candidatos às vagas de administrador de empresas em uma indústria foram as seguintes: 5,0 ; 7,2 ; 6,8 ; 5,3 ; 8,3 ; 7,2 ; 8,0 ; 8,2 ; 6,8 ; 6,3 ; 7,1 ; 8,9 ; 5,9 ; 8,6 ; 6,3 ; 5,3 ; 7,2. Sabendo-se que a nota de corte para selecionar os aprovados é determinada pelo $7^{\underline{o}}$ decil:

a) Determine a nota de corte
b) Quais as notas dos candidatos aprovados?

4. Considere os dois conjuntos de dados amostrais abaixo:

C1={2,53 ; 5,47 ; 2,47 ; 5,57 ; 5,78 ; 6,35 ; 4,34 ; 9,56 ; 8,98 ; 7,34}

C1={16,45 ; 16,09 ; 15,47 ; 16,78 ; 14,98 ; 15,98 ; 14,76 ; 15,23 ; 15,31}

a) Calcule a média, o desvio padrão amostral e a variância amostral para cada conjunto de dados manualmente com o auxílio de uma máquina de calcular e com o Ms. Excel.
b) Qual conjunto apresenta uma maior dispersão entre os dados em relação a sua média? Justifique sua resposta.

5. Dado o conjunto de dados {20, 5, 7, 3, 9, 3, 4, 8, 2}, determine:

   a) O Rol, a média e a mediana.
   b) A moda, o desvio médio e a variância.
   c) O desvio padrão , o quartil Q_3, o decil D_5 e o percentil P_{10}

TEMA 2 – Probabilidade

CONTEÚDO 1 - Conceito e Definição

Podemos demonstrar ao aluno que ele já tem o conceito de probabilidade formado em sua mente, fazendo pra ele uma simples pergunta: Se jogarmos uma moeda qual a probabilidade dar cara? Veremos que eles responderão 50%, ou de uma em duas, ou seja, 1/2. Se perguntarmos qual a probabilidade de sair coroa eles responderão a mesma coisa. Portanto devemos iniciar o nosso estudo de probabilidade com nossos alunos já despertando neles a naturalidade do conceito. É muito importante explicar ao aluno que 50%=50/100=1/2=0,50=0,5, ou seja, cinqüenta por cento, significa dividir 50 por cem, o símbolo %, que dizer exatamente isso, a barra do símbolo (/) significa dividir e os dois zeros, significa dividir por cem (100).

Vamos fazer a seguinte pergunta aos nossos alunos: Se eu jogar uma moeda, alguém pode me dizer com certeza que resultado vai dar? Certamente eles nos responderão que não.

Agora vamos fazer a seguinte pergunta: Vocês podem descrever o conjunto de todas as possibilidades que pode ocorrer no lançamento de uma moeda? Eles nos responderam que pode ocorrer cara (K), ou coroa (C).

A partir daí podemos conceituar experimento aleatório e espaço amostral, pois eles já construíram esses conceitos e é uma coisa natural pra todos nós.

Explicamos para nossos alunos que o lançamento de uma moeda é um experimento aleatório, pois fizemos uma experiência de lançar uma moeda e não sabemos afirmar o que vai ocorrer se vai ocorrer cara (K) ou coroa (C). Por outro lado sabemos descrever o conjunto de todas as possibilidades que pode ocorrer, sabemos que ou vai dar cara (K) ou vai dar coroa (C). Essa é a definição de experimento aleatório.

Ao conjunto de todas as possibilidades de um experimento aleatório, nós chamamos de espaço amostral, e denotamos esse conjunto pele letra grega ômega Ω.

Portanto o espaço amostral do experimento aleatório do lançamento de uma moeda é:

$$\Omega = \{K, C\}$$

A partir desses exemplos e dessas discussões feita na sala de aula, onde muitos já entenderam os conceitos e definições do que desejamos falar, podemos então formalizar essas definições.

Experimento Aleatório.

São chamados de experimentos aleatórios aqueles que repetidos em idênticas condições, produzem resultados que não podem ser previstos com certeza, porém conseguimos descrever o conjunto de todos os resultados possíveis que podem ocorrer.

EXEMPLOS

1) Lançar um dado e observar o número da face de cima.
2) Lançar uma moeda.
3) De um lote de 30 peças defeituosas e 50 boas, retirar 5 peças e observar o número de defeituosas.

Espaço Amostral

Chamamos de espaço amostral o conjunto formado por todos os resultados possíveis de um experimento aleatório. Indicamos por Ω.

Dos experimentos acima temos os seguintes espaços amostrais:

1) Ω = { 1,2,3,4,5,6,}
2) Ω = { K , C } , K = Cara e C = Coroa
3) Ω = { 0,1,2,3,4,5,}

EXEMPLOS

Determinar o espaço amostral dos seguintes experimentos aleatórios

a) Duas moedas são lançadas simultaneamente e observa-se o número de caras.
Ω = {0,1,2}

b) Uma moeda é lançada duas vezes e observa-se a face de cima.
Ω = {(K,K),(K,C),(C,K),(C,C)}

c) Um casal planeja ter 3 filho. Observa-se a seqüência dos sexos
Ω = {(M,M,M),(M,M,F),(M,F,M),(M,F,F),(F,F,F),(F,F,M),(F,M,F),(F,M,M)}

Vamos dar agora o seguinte exemplo em sala de aula: Suponha que eu jogue um dado aqui na sala, alguém pode me descrever o conjunto de possibilidades que pode ocorrer? A turma responderá que ou vai dar o número 1, ou o 2, o 3, o 4, o 5 ou o 6, ou seja, teremos construído o espaço amostral do lançamento de um dado.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

Podemos perguntar agora o seguinte: Se eu jogar o dado, qual a probabilidade de ocorrer o número 5? A turma chegará mais uma vez a uma conclusão que a probabilidade é de uma em seis, ou seja, de $\frac{1}{6} \cong 0{,}1667 = 16{,}67\%$. Pois temos em um dado 6 números e escolhemos um desses seis.

Vamos fazer a seguinte explanação: Suponha que eu irei jogar um dado e o aluno de nossa sala, digamos João, me diga, professor, vai dar um número par, então se eu jogar o dado e ocorrer o número 2, ou 4, ou 6, então terá ocorrido o evento par, como afirmou João. A partir dessa explanação eu defino um evento de um experimento aleatório.

Evento

Seja Ω o espaço amostral de um experimento aleatório. Chamamos de evento todo subconjunto de Ω.

Dizemos que um evento A ocorre se, realizado o experimento o resultado obtido pertence a A

EXEMPLOS

1) Um dado é lançado e observa-se o número da face de cima.
$\Omega = \{1,2,3,4,5,6,\}$

EVENTOS

A: Observar número par. A={2,4,6}
B: Observar número impar B={1,3,5}
C: Observar número menor que quatro. C={1,2,3}

2) Uma moeda é lançada 2 vezes e observa-se a seqüência de caras e coroas.
$\Omega = \{(K,K),(K,C),(C,K),(C,C)\}$

EVENTOS
A: Observar cara no segundo lançamento. A={(K,K),(C,K)}
B: Não observar coroa. B={(K,K)}
C: Observar exatamente uma coroa. C={ (K,C),(C,K)}

Sugiro que o professor quando estiver dando essa aula, ele tenha em mão uma moeda e um dado, para fazer de forma real esses experimentos, isso irá ajudar muito os alunos que têm dificuldades de abstração, além de tornar a aula mais participativa.

Podemos então fazer um experimento aleatório simples, como um lançamento de uma moeda, vamos então conseguir uma moeda e lançar essa moeda, por exemplo, na sala de aula onze vezes e observar o número de caras e coroas.

Eu escolhi um número impar para excluir a possibilidade de dar o mesmo número de caras e coroas, pois o conceito que quero formular agora eu não desejo que o aluno confunda com o que nós já fizemos que foi dizer a probabilidade de dar cara ou coroa no lançamento de uma moeda.

Quando eu estava escrevendo esse texto eu fiz meu experimento e observei o evento cara (K) sete vezes e Coroa (C), 4 vezes. Vamos perguntar para nosso aluno agora, quantos por cento das vezes que lancei a moeda, nós observamos o evento cara?

Nesse momento surgirá uma oportunidade de revermos e aplicarmos o conceito de proporção e de regra de três simples. Podemos então montar uma regra de três simples, onde 11 está para 100%, pois 11 foi o total de vezes que lançamos a moeda e 7 está para x%, pois sete foi o número de vezes que observamos o evento cara>

$$11 \text{ ———————— } 100\%$$
$$7 \text{ ———————— } x\%$$

$$11 \cdot x\% = 7 \cdot 100\% \Rightarrow x\% = \frac{7}{11} \cdot 100\% \Rightarrow x\% \cong 0,6364 \cdot 100\% \Rightarrow x\% \cong 63,64\%$$

Note resolvi a regra de três simples de forma a ficar evidente que para calcularmos quantos por cento corresponde 7 de 11, basta dividirmos 7 por 11, que obteremos a resposta, não precisando desta forma montar regra de três simples. Pois $\frac{7}{10} \cong 0,6364 = 63,64\%$

A cabamos de calcular o percentual das vezes que observamos o evento cara, podemos fazer o mesmo para o evento coroa, ou simplesmente subtrairmos 63,64% de 100% para obtermos o percentual das vezes que observamos o evento coroa, pois só temos esses dois eventos, logo temos que 36,36% das vezes que lançamos a moeda observamos o evento coroa.

A partir dessa prática em sala de aula, informamos ao nosso aluno que calculamos freqüência relativa da ocorrência do evento cara e coroa, em nosso experimento aleatório do lançamento de uma moeda. Podemos então definir formalmente esse conceito.

Freqüência relativa

Em um experimento aleatório, não sabemos qual evento ocorrerá, porém uns ocorrem mais que outros. Queremos associar números a cada evento, que nos dêem uma indicação quantitativa da sua ocorrência. Então definimos freqüência relativa da seguinte maneira:

Definição: Seja $\Omega = \{a_1, a_2, ..., a_k\}$ o espaço amostral de um experimento aleatório qualquer. Suponha que o experimento seja repetido N vezes, nas mesmas condições. Seja n_i o número de vezes que ocorre o evento elementar (unitário) $\{a_i\}$. Definimos a freqüência relativa do evento $\{a_i\}$, sendo o número f_i dado por $f_i = \frac{n_i}{N}$.

Exemplo: Suponha que lançamos um dado 17 vezes e observamos o número 1; três vezes, o número 2; uma vez;o número 3; duas vezes; o número 4; cinco vezes; o número 5; duas vezes e o número 6; quatro vezes então a freqüência relativa destes eventos são:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

$$f_1 = \frac{3}{17} \cong 0,1765 = 17,65\%$$
$$f_2 = \frac{1}{17} \cong 0,0588 = 5,88\%$$
$$f_3 = \frac{2}{17} \cong 0,1176 = 11,76\%$$
$$f_4 = \frac{5}{17} \cong 0,2942 = 29,41\%$$
$$f_5 = \frac{2}{17} \cong 0,1176 = 11,76\%$$
$$f_6 = \frac{4}{17} \cong 0,2353 = 23,53\%$$

Note que cada freqüência relativa é um número entre 0 e 1, ou seja, entre 0% e 100%. Esta é uma propriedade da freqüência relativa.

$$0 \leq f_i \leq 1 \Leftrightarrow 0\% \leq f_i \leq 100\% \text{ ; onde } 1 \leq i \leq 6$$

E que a soma das freqüências relativa da 1, ou seja, 100%. Esta é mais uma propriedade de freqüência relativa de eventos elementares (unitários) de um experimento aleatório.

$$f_1 + f_2 + f_3 + f_4 + f_5 + f_6 = 1 = 100\%$$

Considere o evento A, ocorrer em um lançamento de um dado os números 2 ou 5, ou seja A={2,5}. Temos então que a frequência relativa deste evento, com relação ao exemplo acima é:

$$f_A = \frac{1+2}{17} = \frac{1}{17} + \frac{2}{17} = f_1 + f_2$$

Isto é, a frequência relativa de um evento, é a soma das freqüências relativas dos elementos contidos no evento. Esta é mais uma propriedade de freqüência relativa.

Agora façamos a seguinte pergunta para nossos alunos: Em uma corrida de cavalos, onde 5 cavalos de nomes americano, boitacas, frajola, rabanete e furacão disputam quem vence, qual a probabilidade do cavalo de nome americano ganhar? Certamente eles vão responder de 1 em 5, ou seja, $\frac{1}{5}=0,20=20\%$, pois americano é um dos cinco cavalos que disputam a corrida, e fizemos de forma natural as probabilidades de cada um vencer a corrida sendo a mesma, pois não temos nenhuma informação sobre os cavalos.

Agora vamos dar pro nossos alunos a seguinte informação, em 12 corridas disputadas por esses mesmos cinco cavalos, sabemos que Americano venceu 3; boitacás venceu 1; frajola venceu 4; rabanete venceu 2 e furacão venceu 2.

Vamos fazer pra eles a seguinte pergunta: Em posse dessas informações, se você tivesse que a postar na próxima corrida entre esses cinco cavalos, em que você apostaria? Por quê? Com certeza os alunos vão responder que apostariam no frajola. Porque ele foi quem venceu um maior número de vezes.

Agora se perguntássemos para turma, qual a probabilidade de cada um desses cavalos ganharem a corrida? Eles iriam nos responder:

A probabilidade de americano ganhar é de $\frac{3}{12}=0,25=25\%$

A probabilidade de boitacás ganhar é de $\frac{1}{12}\cong 0,0833=8,33\%$

A probabilidade de frajola ganhar é de $\frac{4}{12}\cong 0,3333=33,33\%$

A probabilidade de rabanete ganhar é de $\frac{2}{12}\cong 0,1667=16,67\%$

A probabilidade de furação ganhar é de $\frac{2}{12}\cong 0,1667=16,67\%$

Eles terão então associado à relação que existe na prática entre probabilidade e freqüência relativa. Podemos agora então formalizar pra eles a definição de probabilidade.

Definição de probabilidade

Considere o espaço amostral $\Omega=\{a_1,a_2,...,a_K\}$. A cada evento $\{a_i\}$, associamos um número real, indicado por $P(a_i)$, que chamamos de probabilidade do evento $\{a_i\}$ ocorrer, que satisfaz as seguintes condições:

1) $0\leq P(a_i)\leq 1,\ \forall i\in\{1,2,...,K\}$.
2) $\sum_{i=1}^{K} P(a_i)=P(a_1)+P(a_2)+...+P(a_k)=1$.

Dizemos que os números $P(a_1)$, $P(a_2)$,...,$P(a_k)$, definem uma probabilidade sobre Ω.

As mesmas propriedades que verificamos para freqüência relativa valem para probabilidade. Entre elas chamamos atenção para as seguintes:

A probabilidade de um evento vazio ($\varnothing$) ocorrer é zero, isto é $P(\varnothing)=0$.

E a outra é a probabilidade de ocorrer o espaço amostral, que é sempre igual a 1, ou seja, 100%. ($P(\Omega)=1$).

Vejamos alguns problemas simples que envolvem probabilidade, que podemos propor aos nossos alunos, onde eles vão utilizar os conceitos de probabilidade e os conceitos já estudados.

Problema 1: Uma moeda é construída utilizando-se para a face cara (K) o chumbo e para face coroa (C) o alumínio, de modo que a probabilidade de ocorrer cara em um lançamento desta moeda é três vezes a probabilidade de sair coroa. Determine a probabilidade de sair cara e coroa no lançamento desta moeda.

Temos que o espaço amostral no lançamento de uma moeda é $\Omega=\{K,C\}$, temos também que $P(\Omega)=1$, ou seja, $P(K)+P(C)=1$. Temos também que a probabilidade de sair cara é três vezes a probabilidade de sair coroa, isto é, $P(K)=3\cdot P(C)$. Logo temos um outro conceito já estudado pelos alunos que é a resolução de sistemas de equações de primeiro grau. Para o aluno visualizar melhor este sistema, podemos fazer $x=P(K)$ e $y=P(C)$, assim temos o sistema abaixo:

$$\begin{cases} x+y=1 \\ x=3\cdot y \end{cases} \Rightarrow 3\cdot y+y=1 \Rightarrow 4\cdot y=1 \Rightarrow y=\frac{1}{4} \Rightarrow x=\frac{3}{4}$$

Logo $P(K)=\frac{3}{4}$ e $P(C)=\frac{1}{4}$

Problema 2: Um dado é viciado de modo que a probabilidade de sair um número par é o dobro da probabilidade de sair um número impar, entretanto a probabilidade de ocorrer números impares são iguais e a probabilidade de ocorrer números pares também são iguais, isto é, $P(1)=P(3)=P(5)$ e $P(2)=P(4)=P(6)$. Determine a probabilidade de ocorrer o evento A={1,2,4}.

Vamos fazer $x=P(1)=P(3)=P(5)$ e $y=P(2)=P(4)=P(6)$, como a probabilidade de sair um número par é o dobro da probabilidade de sair um número impar temos que $y=2\cdot x$, como o espaço amostral de um dado é $\Omega=\{1,2,3,4,5,6\}$ e $P(\Omega)=1$, temos então:

$$P(1)+P(2)+P(3)+P(4)+P(5)+P(6)=1 \Rightarrow x+y+x+y+x+y=1 \Rightarrow$$

$$x+2\cdot x+x+2\cdot x+x+2\cdot x=1 \Rightarrow 9\cdot x=1 \Rightarrow x=\frac{1}{9} \Rightarrow y=\frac{2}{9}$$

Portanto a probabilidade de ocorrer o evento A é:

$$P(A) = P(1) + P(2) + P(4) = \frac{1}{9} + \frac{2}{9} + \frac{2}{9} = \frac{5}{9}$$

É importante que o professor desenvolva no aluno a habilidade de ler, interpretar, entender e formular de forma correta os problemas, uma das maiores dificuldades das pessoas hoje em dia, inclusive de estudantes de graduação, é ler compreender o que o problema pede e saber equacioná-lo de maneira correta, aliás, este processo consiste em 99% da resolução do problema.

O professor deve fazer uso de todos os métodos eficazes de ensino, fazendo uso do que há de melhor neles. Temos o lúdico, o construtivismo, o tradicional, temos as tecnologias, etc. Devemos conhecer e utilizar todas as ferramentas que nos ajudaram no processo de ensino, digo isso porque tem-se abandonado práticas importantes no ensino, a parte de resolver exercícios, resolver problemas em disciplinas quantitativas, como matemática, estatística, física, química, etc, são de fundamental importância no desenvolvimento do raciocínio lógico dos aluno, na compreensão de problemas e no equacionamento correto desses problemas, desenvolvendo principalmente a compreensão do que ele lê, isso é de fundamental importância na vida de uma pessoa.

Portanto neste curso pretendo te dar ferramentas que você professor possa aplicar em sua sala de aula e provocá-lo a experimentar e a experimentar o resultado, fazendo com isso sua própria experiência de ensinar, e que possa divulgar estes resultados e possa transmitir suas práticas e experiências para outros professores com eu e muitos mais que estão dispostos a aprender e sedentos por resultados práticos de aprendizado em nossos alunos.

CONTEÚDO 2 - Probabilidade Condicional

Probabilidade condicional

A probabilidade condicional como o próprio nome já diz, é uma probabilidade condicionada, ou seja, sujeita a uma condição, esta condição é o fato de um determinado evento ter ocorrido, isto é, sabendo-se que um determinado evento ocorreu, deseja-se saber a probabilidade de que um outro evento ocorra.

Por exemplo, sabendo-se que o evento B ocorreu, qual a probabilidade de que o evento A ocorra? Ou podemos dizer, dado que o evento B ocorreu, qual a probabilidade que o evento A ocorra? Ou ainda, qual a probabilidade de ocorrer o evento A, dado (ou sabendo-se) que o evento B ocorreu?

Simbolizamos estas perguntas com a seguinte notação:

$$P(A/B)$$

Vejamos um exemplo para ficar mais claro.

Exemplo: Considere o espaço amostral do lançamento de um dado $\Omega = \{1,2,3,4,5,6\}$, seja o evento $A = \{1,3\}$ e o evento $B = \{3,4,6\}$. Suponha que o dado seja jogado e seja informado que o evento *B* ocorreu, pede-se para determinar a probabilidade de que o evento *A* também tenha ocorrido.

Note que se ocorreu o evento B, sabemos que ou ocorreu o número 3, ou o 4, ou o 6, para que o evento A também tenha ocorrido, só temos uma possibilidade entre essas três, que é justamente ter ocorrido o número 3, pois esse número é o único dos três que está contido em A, logo essa probabilidade é de $\frac{1}{3}$. E escrevemos como abaixo:

$$P(A/B) = \frac{1}{3}$$

Note que podemos deduzir a partir desse exemplo uma relação para o cálculo da probabilidade condicional. A interseção entre A e B é $A \cap B = \{3\}$ e a probabilidade de ocorrer a interseção em um lançamento de um dado é: $P(A \cap B) = \frac{1}{6}$, e a probabilidade de ocorreu o evento B é: $P(B) = \frac{3}{6}$, logo podemos obter $P(A/B)$, da seguinte maneira:

$$P(A/B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{6}}{\frac{3}{6}} = \frac{1}{6} \cdot \frac{6}{3} = \frac{1}{3}$$

Agora vamos supor o contrário, isto é, suponha que o evento A tenha ocorrido, desejamos agora saber a probabilidade de que o evento B ocorra.

Se ocorreu o evento A, é por que no lançamento do dado ou ocorreu o número 1 ou o número 3, logo para que o Evento B ocorra a única possibilidade é ocorrer o número 3, pois o número 3, está contido em B, logo essa probabilidade é de uma em duas, ou seja, $\frac{1}{2}$.

Podemos calcular essa probabilidade com a relação acima descrita

$$P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{\frac{1}{6}}{\frac{2}{6}} = \frac{1}{2}$$

Portanto temos uma relação de fundamental importância no cálculo de probabilidade que é dada abaixo::

$$P(A/B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B/A) = \frac{P(B \cap A)}{P(A)}$$

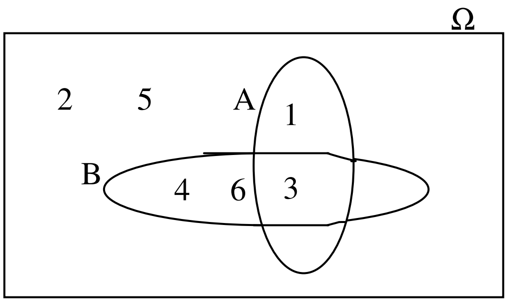
Uma outra forma de entendermos o cálculo da probabilidade condicional é calcularmos P(A/B) usando B como o novo espaço amostral reduzido, dentro do qual queremos calcular a probabilidade do evento A.

A utilização do diagrama de Venn facilita muito o entendimento do aluno, tudo que for possível representar geometricamente, vai facilitar a visualização do aluno

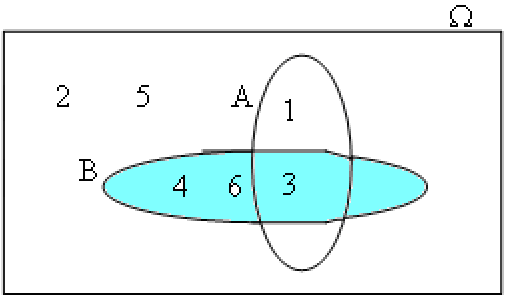
Considere o nosso exemplo acima, onde $\Omega = \{1, 2, 3, 4, 5, 6\}$; $A = \{1, 3\}$ e $B = \{3, 4, 6\}$.

Para calcularmos a probabilidade do evento P(A/B)

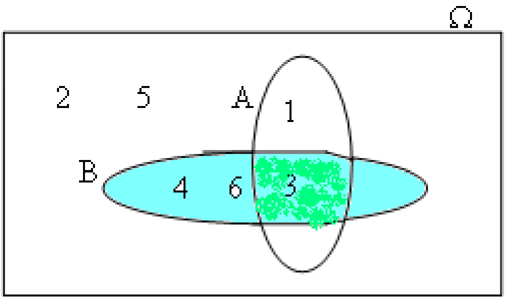
Vejamos então o diagrama de Venn do nosso exemplo:



Se ocorreu o evento B, devemos focar nele, ele é o nosso espaço amostral reduzido, esquecemos os demais números, que no caso são o 1, o 2 e o 5. No diagrama abaixo nós colorimos o evento que ocorreu, que foi o B, para focarmos nele.



Já que o evento B ocorreu, temos que a probabilidade de que A ocorra é de um número, que é justamente o 3, de um total de três números, que são o 3, o 4 e o 6, pois apenas um número, que é o 3, faz parte da interseção entre os conluntos, logo P(A/B)=1/3. A interseção entre os conjuntos A e B está destacada no diagrama abaixo



Vejamos mais um exemplo de cálculo de probabilidade condicional.

Exemplo: Foram selecionados 400 pessoas em uma cidade, cujo sexo e estado civil estão na tabela abaixo:

	Solteiros (S)	Casados (C)	Desquitados (D)	Viuvos (V)	Total
Masculino (M)	50	60	40	30	180
Feminino (F)	150	40	10	20	220
Total	200	100	50	50	400

Supondo que escolhemos desse grupo de forma aleatória uma pessoa do sexo masculino, qual a probabilidade dessa pessoa ser solteira? Ou em outras palavras, qual a probabilidade de uma pessoa desse grupo ser solteira, se sabemos que ela é do sexo masculino? Queremos então determinar P(S/M), pois ocorreu neste caso o evento masculinas (M), e desejamos determinar a probabilidade de que ocorra o evento solteiro (S). Logo:

P(S/M)=50/180=5/18, pois é a probabilidade de ocorrer solteiro, dado que ocorreu o evento masculino. Logo o evento masculino se torna o espaço amostral reduzido, onde calcularemos a probabilidade de o evento solteiro.

Se escolhermos uma pessoa desse grupo e ela é solteira, qual a probabilidade dessa pessoa ser do sexo masculino? Ou em outras palavras, qual a probabilidade de uma pessoa desse grupo ser do sexo masculino, se sabemos que essa pessoa é solteira? O que queremos calcular é P(M/S), pois ocorreu o evento solteiro (S) e queremos determinar o evento masculino (M). Logo:

P(M/S)=50/200=1/4=0,25=25%, pois é a probabilidade de ocorrer masculino, dado que ocorreu o evento solteiro. Logo o evento solteiro se torna o espaço amostral reduzido, onde calcularemos a probabilidade de o evento masculino. Podemos observar que P(S/M)≠P(M/S).

P(F/D)=10/50=1/5=0,20=20%
P(D/F)=10/220=1/22.
P(V/M)=30/180=1/6
P(M/V)=30/50=3/5=0,60=60%.

Podemos calcular essas probabilidades usando as relações deduzidas

P(S/M)= $\dfrac{P(S \cap M)}{P(M)} = \dfrac{\frac{50}{400}}{\frac{180}{400}} = \dfrac{50}{400} \cdot \dfrac{400}{180} = \dfrac{50}{180} = \dfrac{5}{18}$.

P(M/S)= $\dfrac{P(M \cap S)}{P(S)} = \dfrac{\frac{50}{400}}{\frac{200}{400}} = \dfrac{50}{400} \cdot \dfrac{400}{200} = \dfrac{50}{200} = \dfrac{1}{4} = 0{,}20 = 20\%$.

Vamos ver mais um exemplo de probabilidade condicional.

Exemplo: Considere o experimento aleatório do lançamento de dois dados D1 e D2, onde observamos os números do dado D1 e do dado D2, representado pelo par ordenado (d1,d2). Logo este experimento tem o seguinte espaço amostral Ω:

$$\Omega = \left\{ \begin{array}{l} (1,1);\ (1,2);\ (1,3);\ (1,4);\ (1,5); (1,6) \\ (2,1); (2,2); (2,3); (2,4); (2,5); (2,6) \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ (6,1); (6,2); (6,3); (6,4); (6,5); (6,6) \end{array} \right\}$$. Considere os seguintes eventos:

A: Observar o número 4 em D2.

B: A soma dos números de D1 e D2 é 5.
Determinar P(A/B) e P(B/A).
A={(1,4);(2,4);(3,4);(4,4);(5,4);(6,4)}.
B={(1,4);(2,3);(3,2);(4,1)}.

Logo P(A/B)=$\frac{P(A\cap B)}{P(B)}=\frac{\frac{1}{36}}{\frac{4}{36}}=\frac{1}{36}.\frac{36}{4}=\frac{1}{4}=0{,}25=25\%$

e P(B/A)=$\frac{P(B\cap A)}{P(A)}=\frac{\frac{1}{36}}{\frac{6}{36}}=\frac{1}{36}.\frac{36}{6}=\frac{1}{6}.$

CONTEÚDO 3 - Teorema da Multiplicação e Teorema da Probabilidade Total

Teorema da multiplicação

Temos da definição de probabilidade condicional, que $P(A/B)=\frac{P(A\cap B)}{P(B)}$ e $P(B/A)=\frac{P(B\cap A)}{P(A)}$. Logo P(A∩B)=P(B).P(A/B) e P(A∩B)=P(A).P(B/A), isto é, a probabilidade de ocorrer os eventos A e B é igual ao produto da probabilidade de ocorrer um deles pela probabilidade condicional do outro dado que o primeiro ocorreu.

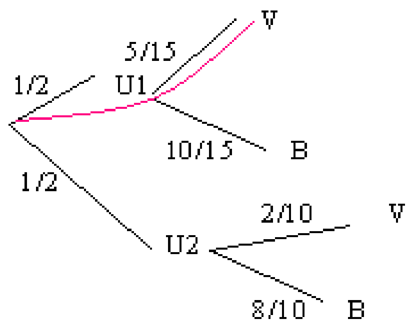
Mas veja que para se aplicar este teorema, não precisamos usar nenhuma relação matemática, podemos apenas construir um diagrama de árvore, para entendermos sua aplicação. Vejamos o exemplo abaixo:

Exemplo: Em uma sala encontra-se duas urnas, a urna U1 contém 5 bolas vermelhas e 10 bolas brancas, a urna U2 contém 2 bolas vermelhas e 8 brancas. Uma pessoa vai entrar nessa sala, escolher aleatoriamente uma das urnas e tirar também aleatoriamente uma bola desta urna, determine a probabilidade dessa pessoa escolher a urna 1 e retirar dessa urna uma bola vermelha.

Note que desejamos saber a probabilidade de ocorrer o evento U1 e o evento bola vermelha, ou seja, a interseção dos eventos. Podemos resolver este problema utilizando a relação descrita acima, veja a seguir:

P(U1∩V) = P(U1).P(V/U1) =$\frac{1}{2}.\frac{5}{15}=\frac{5}{30}=\frac{1}{6}$, pois P(U1)=1/2, que é a probabilidade da pessoa escolher a urna U1 e P(V/U1)=5/15, que é a probabilidade da Pessoa escolher uma bola vermelha, dado que ela escolheu a urna U1..

Agora vejamos a construção completa do diagrama de árbore desse problema.



Note que a probabilidade de ocorrer o evento urna U1 e bola vermelha e o produto das probabilidades dos galhos da árvore que nos leva a urna U1 e a bola vermelha, este caminho está evidenciado pela linha vermelha no diagrama de árbore acima, logo

$$P(U1 \cap V) = \frac{1}{2} \cdot \frac{5}{15} = \frac{1}{6}$$

De igual forma obtemos outras probabilidades:

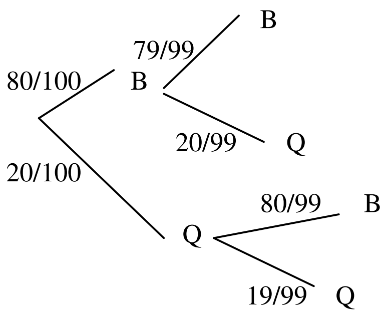
P(U1∩B)= $\frac{1}{2} \cdot \frac{10}{15} = \frac{10}{30} = \frac{1}{3}$

P(U2∩V)= $\frac{1}{2} \cdot \frac{2}{10} = \frac{2}{20} = \frac{1}{10} = 0{,}10 = 10\%$

P(U2∩B)= $\frac{1}{2} \cdot \frac{8}{10} = \frac{8}{20} = \frac{2}{5} = 0{,}40 = 40\%.$

Vejamos um outro ejemplo em que utilizamos o diagrama de árbore, para um problema que envolve amostra sem reposição.

Exemplo: Em um lote de 100 lâmpadas, existem 80 boas (B) e 20 queimadas (Q). Uma lâmpada é escolhida ao acaso e, sem reposição desta, outra é escolhida também ao acaso. Determine a probabilidade escolhermos a primeira lâmpada boa e a segunda queimada.

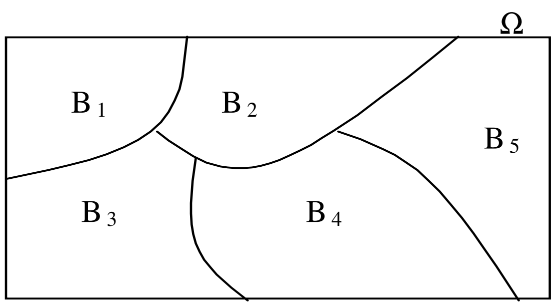


A probabilidade de escolhermos uma lâmpada boa e outra queimada sem reposição, seria o produto das probabilidades que estão nos ramos do caminho boa (B) e queimada (Q). Logo temos a probabilidade $\frac{80}{100}.\frac{20}{99}=\frac{4}{5}.\frac{20}{99}\cong 0{,}16=16\%$. Notamos que a probabilidade de escolhermos uma queimada (Q) e uma boa (B) têm a mesma probabilidade de escolhermos uma boa (B) e uma queimada (Q). De fato $\frac{20}{100}.\frac{80}{99}=\frac{1}{5}.\frac{80}{99}\cong 0{,}16=16\%$. A probabilidade de escolhermos duas boas é $\frac{80}{100}.\frac{79}{99}=\frac{4}{5}.\frac{79}{99}\cong 0{,}64=64\%$ e a probabilidade de escolhermos duas queimadas é $\frac{20}{100}.\frac{19}{99}=\frac{1}{5}.\frac{19}{99}\cong 0{,}04=4\%$.

Teorema da probabilidade total

Antes de formalizarmos este teorema vamos formalizar o conceito de uma partição de um espaço amostral, ou seja, o conceito de partição de um conjunto, que no nosso caso é o espaço amostral, que aqui assume o lugar do conjunto universo na teoria de conjuntos.

Vejamos op diagrama de Venn abaixo:



Visualize o diagrama acima como uma vidro retangular, que é o espaço amostral Ω, onde este vidro tomou uma pancada e rachou em cinco pedaços, que no diagrama estão representados pelos eventos B_1 ,B_2 ,B_3, B_4 e B_5, estes eventos formam uma partição do espaço amostral Ω. Note que cada evento desses é não vazio e não existe interseção entre dois deles quaisquer, e a união deles resulta em todo o espaço amostral, está é a condição para que eventos formem uma partição de um espaço amostral, não importa quantos eventos formam a partição. Abaixo formalizamos a definição de uma partição de um espaço amostral.

Sejam $B_1,B_2,...,B_n$ n eventos do espaço amostral Ω, dizemos que eles formam uma partição de Ω se:

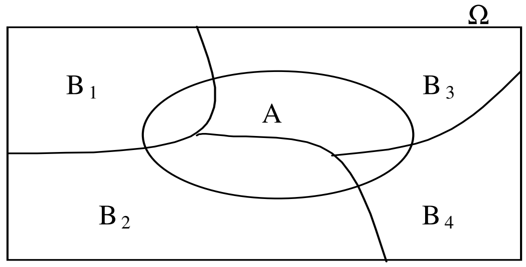
1) $P(B_i) > 0$.

2) $B_i \cap B_j = \varnothing$, se $i \neq j$.

3) $\bigcup_{i=1}^{n} B_i = \Omega$.

Isto é, os $B_{i's}$ são dois a dois mutuamente exclusivos e exaustivos.

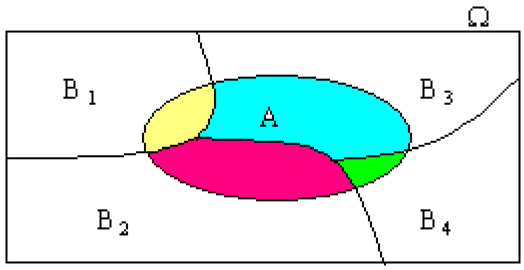
Agora considere o diagrama abaixo, onde temos uma partição do espaço amostral, formada por quatro eventos, B_1 ,B_2 ,B_3, B_4 e temos um evento A qualquer, também deste espaço amostral



Note que podemos escrever o evento A da seguinte maneira:

A= $(B_1 \cap A) \cup (B_2 \cap A) \cup .(B_3 \cap A) \cup (B_4 \cap A)$

Ou seja, sendo a união das interseção de $(B_1 \cap A)$, $(B_2 \cap A)$, $(B_3 \cap A)$ e $(B_4 \cap A)$, que estão representadas nas cores, amarela, vermelha, azul e verde respectivamente no diagrama abaixo:



Como os conjuntos $(B_1 \cap A)$, $(B_2 \cap A)$, $(B_3 \cap A)$ e $(B_4 \cap A)$ são dois a dois mutuamente exclusivos, isto é, a interseção entre dois quaisquer destes conjuntos é o conjunto vazio, temos que

$$P(A)=P(B_1 \cap A)+P(B_2 \cap A)+P(B_3 \cap A)+P(B_4 \cap A).$$

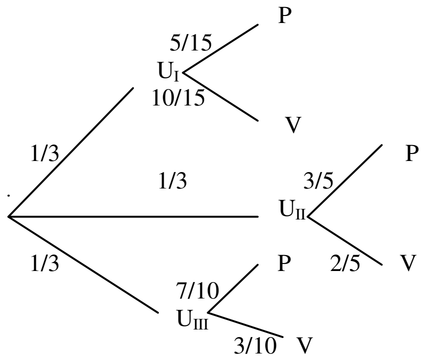
Este resultado é conhecido como teorema da probabilidade total e generalizamos para uma partição qualquer de n eventos, veja abaixo.

$$P(A)=P(B_1 \cap A)+P(B_2 \cap A)+...+P(B_n \cap A).$$

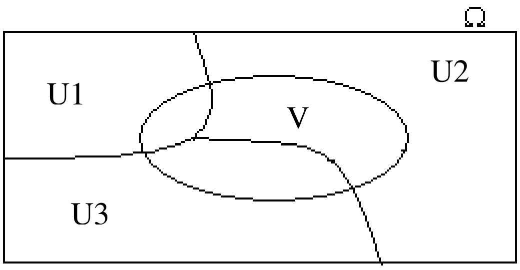
Vejamos um exemplo de aplicação deste teorema em um problema.

Exemplo: Uma urna I (U_I) tem 5 bolas pretas e 10 vermelhas; outra urna II (U_{II}) tem 3 bolas pretas e 2 vermelhas, e a urna III (U_{III}) tem 7 bolas pretas e 3 vermelhas. Uma urna é selecionada ao acaso e dela também aleatoriamente é retirada uma bola. Qual a probabilidade desta bola ser vermelha?

Vamos primeiramente construir o diagrama de árvore de toda situação que temos:



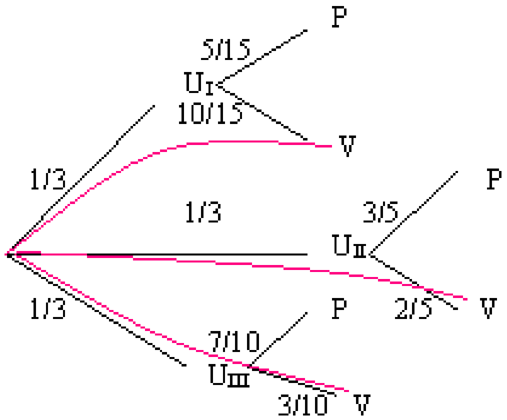
Temos que as urnas U_I, U_{II} e U_{III}, formam uma partição do espaço amostral, pois o espaço amostral é composto pelo total de bolas e essas bolas estão divididas em três urnas e note que não há interseção entre as urnas, pois a bola que está na urna U1 não está na urna U2 e nem na U3, e assim é para todas as urnas.
Note que estamos interessados em saber qual a probabilidade de ocorrer o evento bola vermelha, não importando de que urna esta bola tenha vindo, veja o problema que temos no diagrama de Venn abaixo:



Logo usamos o teorema da probabilidade total para resolvermos, onde temos:

P(V)=P($U_1 \cap V$)+P($U_2 \cap V$)+P($U_3 \cap V$)= $\frac{1}{3}.\frac{10}{15}+\frac{1}{3}.\frac{2}{5}+\frac{1}{3}.\frac{3}{10}=\frac{2}{9}+\frac{2}{15}+\frac{1}{10}\cong 0{,}46=46\%$

Este problema pode visualizado apenas utilizando o diagrama de árvore, onde a probabilidade de ocorrer o evento vermelho deve ser calculada levando em consideração todos os caminhos que nos leva ao evento vermelho, esse caminhos estão indicados pelas linhas vermelhas no diagrama abaixo:



logo o conjunto vermelho podemos escrever da seguinte forma: $V=(U_I \cap V)\cup(U_{II}\cap V)\cup(U_{III}\cap V)$, logo a probabilidade de retirarmos uma bola vermelha será $P(V)=P(U_I\cap V)+P(U_{II}\cap V)+P(U_{III}\cap V)=\frac{1}{3}.\frac{10}{15}+\frac{1}{3}.\frac{2}{5}+\frac{1}{3}.\frac{3}{10}=\frac{2}{9}+\frac{2}{15}+\frac{1}{10}\cong 0{,}46=46\%$.

CONTEÚDO 4- Eventos Independentes, Arranjos e Combinação.

Independência de dois eventos

Antes de formalizarmos a definição de eventos independentes, vamos falar sobre eles de uma maneira bem natural, que é falar sobre independência.

Um evento é independente de um outro, se a ocorrência de um deles não altera em nada a probabilidade do outro ocorrer. Por exemplo, se ocorrer o evento *João passar no vestibular*, o fato desse evento ter ocorrido não altera em nada o evento *Maria ir fazer compras*. E vice versa, ou seja, o fato de *Maria ir fazer compras*, não altera em nada a probabilidfade do evento *João passar no vestibular*.

È importante desenvolver no aluno desde as séries iniciais as noções de álgebra e de determinadas demonstrações simples, que o aluno possa acompanhar, por essa razão fazemos umas demonstrações simples abaixo para formalizarmos o conceito de eventos independentes.

Dados dois eventos A e B de um espaço amostral Ω, dizemos que o evento A independe do evento B se, $P(A/B)=P(A)$. Isto é, A independe de B se a ocorrência de B não afeta a probabilidade de A. Notamos que se A independe de B, então B independe de A. Vejamos:

Se A independe de B, então P(A/B)= $\frac{P(A \cap B)}{P(B)}$ =P(A), logo P(A∩B)=P(A).P(B). Como P(B/A)= $\frac{P(A \cap B)}{P(A)} = \frac{P(A).P(B)}{P(A)}$ =P(B), logo B independe de A.

Temos que P(A/B)= $\frac{P(A \cap B)}{P(B)}$, logo P(A∩B)=P(A/B).P(B), se A independe de B temos que P(A/B)=P(A), então P(A∩B)=P(A).P(B).

Podemos então dizer que dois eventos A e B de um espaço amostral Ω, são independentes se P(A∩B)=P(A).P(B).

Exemplo: Uma moeda é lançada três vezes. Logo temos que Ω={(K,K,K),(K,K,C),(K,C,K),(K,C,C),(C,C,C),(C,C,K),(C,K,C),(C,K,K)}.

Considere os eventos:
A: Ocorrerem resultados iguais nos três lançamentos.
B: Ocorrerem pelo menos duas caras.

A={(K,K,K),(C,C,C)} e B={(K,K,C),(K,C,K),(C,K,K),(K,K,K)}. Temos que P(A)=2/8=1/4 e P(B)=4/8=1/2. Como A∩B={(K,K,K)}, temos que P(A∩B)=1/8. Logo temos que P(A∩B)=P(A).P(B), pois 1/8=1/4.1/2. Temos que neste caso os eventos A e B são independentes.

Se dois eventos A e B de um espaço amostral Ω não são independentes, então eles são ditos dependentes.

Exemplo: Dois candidatos A e B disputam o mesmo vestibular para um determinado curso de uma faculdade. A probabilidade do candidato A passar é de 1/2, e do candidato B passar é 3/4. Determine:

a) A probabilidade de ambos passarem no vestibular.
b) A probabilidade de pelo menos um passar no vestibular.

a) Neste caso temos que o evento do candidato A passar no vestibular, independe do evento do candidato B passar no vestibular. Logo A e B são independentes, então P(A∩B)=P(A).P(B)=1/2.3/4=3/8.
b) P(A∪B)=P(A)+P(B)-P(A∩B)=1/2+3/4-3/8=7/8.

Exercício: Mostrar que se dois eventos A e B de um espaço amostral Ω são independentes, então: A e B^C são independentes, A^C e B são independentes e A^C e B^C são independentes.

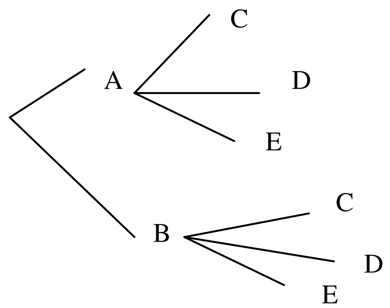
Exercício: Considere o exemplo acima. Determine as seguintes probabilidades:
a) Nenhum dos candidatos passem no vestibular.
b) O candidato A passe no vestibular e o candidato B não.
c) O candidato B passe no vestibular e o candidato A não.
Resp: a) 1/8 b) 1/8 c) 3/8

Métodos de enumeração

- Regra da multiplicação

Suponha que uma primeira decisão possa ser tomada de N maneiras e uma segunda decisão de
M maneiras, então o número de decisões possíveis é M.N.

Exemplo: Suponha que uma primeira decisão possa ser tomada de duas maneiras A e B, e uma segunda decisão possa ser tomada de três maneiras C, D e E.



Logo as decisões possíveis são {{A,C},{A,D},{A,E},{B,C},{B,D},{B,E}}, logo o número de decisões possíveis são 2.3=6.

Permutações

Se temos três objetos A,B e C, nós podemos colocá-los juntos nas seguintes ordens: ABC, ACB , BAC , BCA , CAB , CBA. Logo temos seis maneiras diferentes de arrumar estes três objetos. Se tivermos N objetos, nós teremos N possibilidades (objetos) para a primeira posição, N-1 possibilidades (objetos) para a segunda, N-2 possibilidades (objetos) para a terceira e assim sucessivamente, até a posição N onde teremos apenas uma possibilidade (objeto).

N	N-1	N-2	.	.	.	2	1

Pela regra da multiplicação temos um total de N.(N-1).(N-2)...2.1 possibilidades de permutarmos estes objetos. Se N é um inteiro positivo definimos o fatorial de N sendo N!=N.(N-1).(N-2)...2.1. Definimos 0!=1.

O número de maneira de permutarmos N objetos é dado por P_N= N!.

Arranjo

- Sem Repetição

Suponha que temos n elementos distintos e escolhemos r elementos entre eles. Chamamos de arranjo dos n elementos tomados r a r a qualquer seqüência de r elementos tomados entre estes n elementos, todos distintos.

O número de arranjos de n elementos tomados r a r é denotada por $A_{n,r}$. Para a primeira posição temos n possibilidades, para segunda n-1, e assim sucessivamente até e posição r, onde teremos n-(r-1) possibilidades.

n	n-1	n-2	.	.	.	n-(r-2)	n-(r-1)

Temos que $A_{n,r}$ = n.(n-1).(n-2)...(n-(r-1)) = $\frac{n!}{(n-r)!}$. Pois $\frac{n!}{(n-r)!}=\frac{n.(n-1)...(n-(r-1)).(n-r).(n-(r+1))...2.1}{(n-r).(n-(r+1)...2.1}=n.(n-1).(n-2)...(n-(r-1))$. Logo $A_{n,r}=\frac{n!}{(n-r)!}$.

Exemplo: Sejam a e b dois elementos, logo $A_{2,2}$=2!/(2-2)!=2!=2. As possibilidades são {(a,b),(b,a)}

Exemplo: Sejam a,b e c três elementos, logo $A_{3,3}$=3!/(3-3)!=3!=6. As possibilidades são {(a,b,c),(a,c,b),(b,a,c),(b,c,a),(c,a,b),(c,b,a)}.

Exemplo: Sejam a,b,c e d quatro elementos, logo $A_{4,2}$=4!/(4-2)!=4!/2!=24/2=12. As possibilidades são {(a,b),(b,a),(a,c),(c,a),(a,d),(d,a),(b,c),(c,b)(b,d),(d,b),(c,d),(d,c)}.

Exemplo: Considere agora o número de arranjos deste 3 elementos tomados 2 a 2.

Temos $A_{3,2}$=3!/(3-2)!=3!=6. As possibilidades são {(a,b),(b,a),(a,c),(c,a),(c,b),(b,c)}. Note que não precisamos de uma fórmula matemática para calcularmos arranjos, veja que podemos usar a regra da multiplicação mencionada acima. Entenda que o arranjo de 3 elementos tomados dois a dois, significa em contar quantos arranjos (seqüências) de dois elementos eu posso formar, tomando 2 elementos entre três elementos, logo temos três elementos, digamos {a,b,c}, e queremos contar quantos arranjos (seqüências) podemos formar tomando 2 entre esses três elementos, logo temos duas vagas que estão representadas pelas linhas abaixo:

$$\_\_ \quad \_\_$$

Quantas possibilidades temos para iniciar nossa seqüência? Temos três possibilidades, pois poderíamos iniciar com os elementos a, b ou c, para a segunda posição da seqüência temos duas possibilidades, pois já usamos uma das três para a primeira posição da seqüência, logo temos o produto abaixo:

$$\underline{3}\cdot\underline{2}=6$$

Esse resultado é exatamente do arranjo de três elementos tomados dois a dois.

Quantas possibilidades temos de formar placas de carros, com todas as letras distintas e todos os números distintos? A placa de um carro é formada por três letras e quatro números, coloquemos essa seqüência em espaços vagos abaixo:

— — —   — — — —

Temos os três primeiros espaços para as letras e os quatro últimos para os números, para as letras temos 26 possibilidades, pois são 26 letras que podemos usar, pois incluem o K, W e Y e para os númeors são 10 possibilidades, de 0 a 9, logo temos:

$$\underline{26}\cdot\underline{25}\cdot\underline{24}\ \cdot\ \underline{10}\cdot\underline{9}\cdot\underline{8}\cdot\underline{7} = 78.624.000$$

Logo podemos emplacar 78.624.000 (setenta e oito milhões, seiscentos e vinte e quatro mil) carros, sem repetir nenhuma letra nem números, isso é um exemplo de contagem em que não fazemos uso de fórmulas, este é um problema de arranjo sem repetição.

- Com Repetição

Agora desejamos saber quantos carros podemos emplacar usando três letras e quatro números? Este é um exemplo de arranjo com repetição. Neste casso podemos repetir as letras e os números então temos:

$$\underline{26}\cdot\underline{26}\cdot\underline{26}\ \cdot\ \underline{10}\cdot\underline{10}\cdot\underline{10}\cdot\underline{10} = 175.760.000$$

Combinações

Dado um conjunto M com n elementos, M={a_1,a_2,...a_n}, nós chamamos de combinação dos n elementos, tomados r a r, a todo subconjunto de r elementos do conjunto M.

Exemplo: Considere o conjunto {a,b,c,d}, queremos tomar dois dentre estes quatro elementos. De quantas maneiras podemos fazer isto? Temos as seguintes possibilidades: {a,b},{a,c},{a,d},{b,c},{b,d},{c,d}.

O número de combinações possíveis que podemos fazer com n elementos tomados r a r é indicado por $C_{n,r}$ ou $\begin{pmatrix} n \\ r \end{pmatrix}$, onde $0 \leq r \leq n$. Sendo que $C_{n,r} = \frac{n!}{r!.(n-r)!}$. Considerando o exemplo acima temos que $C_{4,2} = \frac{4!}{2!.(4-2)!} = \frac{24}{2.2} = \frac{24}{4} = 6$.

Exemplo: Dentre 6 pessoas desejamos fazer grupos de dois, quantas são essas possibilidades?

$C_{6,2}=\frac{6!}{2!.(6-2)!}=\frac{720}{2.24}=\frac{720}{48}=15$ possibilidades.

Exemplo: Dentre 6 sons diferentes de um aparelho eletrônico, quantos sinais podemos produzir usando dois sons entre esses 6?

$A_{6,2}=\frac{6!}{(6-2)!}=\frac{720}{24}=30$ possibilidades.]

Exemplo: Dentre um grupo de 6 pessoas temos 4 homens e duas mulheres, queremos formar um grupo de 2 pessoas sendo que 1 é homem e 1 é mulher. Quantas são as possibilidades de formarmos este grupo?

$C_{4,1}=\frac{4!}{1!.(4-1)!}=\frac{24}{6}=4$, $C_{2,1}=\frac{2!}{1!.(2-1)!}=\frac{2}{1}=2$, logo as possibilidades são $C_{4,1}.C_{2,1}=4.2=8$.

Exemplo: Se temos 10 pessoas, sendo 6 homens e 4 mulheres, e queremos formar um grupo de 5 pessoas, sendo 3 homens e 2 mulheres, quantas possibilidades temos para formar este grupo?

$C_{6,3}=\frac{6!}{3!.(6-3)!}=\frac{720}{36}=20$, $C_{4,2}=\frac{4!}{2!.(4-2)!}=\frac{24}{4}=6$, logo as possibilidades são $C_{6,3}.C_{4,2}=20.6=120$.

Exercício: De um determinado grupo de 101 empresas alimentícias que fabricam pelo menos um dos produtos mencionados, sabe-se que 66 fabricam geléias, 62 fabricam sorvetes e 56 produzem chocolates. Destas, 39 fabricam sorvetes e geléias, 42 fabricam sorvetes e chocolates, 38 fabricam chocolates e geléias. Qual a probabilidade de uma empresa escolhida ao acaso fabricar: a) Somente geléia? b) Somente chocolate ou sorvete? c) Chocolate e sorvete e geléia? d) Geléia mais não sorvete? e) Chocolate ou sorvete mas não geléia?
Resp: a) 25/101 , b) 29/101 , c) 36/101 , d) 27/101 , e) 35/101

Exercício: De um grupo de 126 estudantes temos que 66 falam inglês, 52 falam francês e 11 não falam nenhuma destas línguas.

a) I e F são coletivamente exaustivos? Por quê? b) I e F são mutuamente excludentes? por quê? c) Determine $P(I \cap F)$. d) Determine $P(I \cup F)$. e) Qual a probabilidade de escolhermos um estudante que fale apenas inglês? f) Qual a probabilidade de escolhermos um estudante que fale apenas francês?

Resp: c) 3/126 d) 115/126 e) 63/126 f) 49/126.

Exercício: Um cliente de uma loja de roupas e sapatos deseja comprar 4 camisas, 3 calças e 2 sapatos, ele está em duvida entre 6 camisas, 4 calças e 3 sapatos. De quantas maneiras ele pode efetuar esta compra?
Resp: $C_{6,4}.C_{4,3}.C_{3,2}$ = 180 maneiras.

Exercício: Uma dona de casa tem 12 livros em sua estante, sendo 5 matemática, 4 de história e 3 física. a) De quantas maneiras a dona de casa pode arrumar os livros na estante sendo que os livros de história fiquem todos juntos? b) De quantas maneiras a dona de casa pode arrumar os

livros na estante sendo que os livros de matemática e física fiquem todos juntos?
Resp: a) $A_{4,4}.A_{9,9} = 4!.9!$ b) $A_{5,5}.A_{3,3}.A_{6,6} = 5!.3!.6!$.

Exercício: Um rapaz construiu uma rifa onde o prêmio era seu carro. Ele criou cartões com cinco cores diferentes, azul,vermelho, verde, amarelo e branco, os cartões contém 5 dezenas escolhidas entre 10 dezenas. Qual a probabilidade de uma pessoa ganhar o prêmio adquirindo um cartão, se no dia marcado sorteia-se uma cor e cinco dezenas entre as dez?
Resp: $1/C_{5,1}.C_{10,5} = 1/1260$.

Exercício: A caixa econômica federal planeja lançar um jogo onde o jogador deve marcar 6 dezenas das 12 possíveis do cartão. Ele ganha o prêmio se conseguir acertar pelo menos 5 das 6 dezenas. Qual a probabilidade dele ganhar o jogo?
Resp: $(C_{6,5}+C_{6,6})/C_{12,6}=7/924$.

Exercício: Nove livros são colocados ao acaso numa estante. Qual a probabilidade de que 3 livros determinados fiquem juntos?
Resp: $(A_{3,3}.A_{7,7})/A_{9,9} = 1/12$.

Exercício: Em um dos novos jogos comercializados pela caixa econômica federal, o apostador deve marcar 50 dezenas de um cartão contendo 100 dezenas. Posteriormente são sorteadas 20 dezenas. Calcule a probabilidade do apostador ganhar: a) acertando todas as 20 dezenas; b) acertando 16, 17, 18, 19, ou 20 dezenas ou errando as 20 dezenas.
Resp: a) $C_{50,20}/C_{100,20}$ b) $(C_{50,16}+C_{50,17}+C_{50,18}+C_{50,19}+C_{50,20}+C_{50,20})/C_{100,20}$.

Atividade Complementar

1. Um grupo de pessoas foi classificado segundo o peso e o nível de colesterol no sangue, onde a proporção encontra-se na tabela abaixo:

	Peso			
Colesterol	Excesso	Normal	Baixo	Total
Alto	0,20	0,18	0,17	0,55
Normal	0,12	0,10	0,23	0,45
Total	0,32	0,28	0,40	1,00

a) Qual a probabilidade de uma pessoa escolhida ao acaso ter colesterol normal?
b) Se escolhermos uma pessoa desse grupo e ela tiver colesterol alto, qual a probabilidade que ela tenha excesso de peso?

2. Em uma fábrica temos duas máquinas, máquina A e máquina B. A máquina A é responsável por 53,7% da produção, e a máquina B por 46,3% da produção. A máquina A produz 1,3% de peças com defeito e a máquina B 2,5% de peças com defeito.

a) Se uma peça é escolhida ao acaso e verificamos que ela é defeituosa, qual a probabilidade dela ter sido produzida pela máquina B?
b) Se escolhemos uma peça ao acaso, qual a probabilidade dela ser boa?

3. Uma moeda é viciada de modo que a probabilidade de sair K (cara) é 7/3 da probabilidade de sair C (coroa). Se lançarmos essa moeda qual a probabilidade de sair:

a) Cara (K)
b) Coroa (C))

4. Em uma urna temos 20 bolas enumeradas de 1 a 20.

a) Se retirarmos uma bola par dessa urna, qual a probabilidade dessa bola ser maior ou igual a 13?
b) Qual a probabilidade de uma bola menor que 17 ser retirada dessa urna?
c) Qual a probabilidade de uma bola que seja múltipla de 3 e múltipla de 2 ser retirada dessa urna?

5. Em uma urna temos quatro moedas. A moeda M1 é uma moeda normal, a moeda M2 é viciada de tal modo que sair cara (K) é 1219 vezes mais provável que sair coroa (C)), a moeda M3 tem duas caras (K) e a moeda M4 tem duas coroas (C). Uma moeda é escolhida ao acaso e lançada.

a) Se o resultado obtido foi coroa (C), qual a probabilidade da moeda lançada ter sido a moeda M2?
b) Qual a probabilidade de observarmos moeda M3 e cara (K)?
c) Qual a probabilidade de observarmos coroa (C)?

Glossário:

Amostra: Uma amostra é um subconjunto de indivíduos da população alvo. Existem dois tipos de amostras, as probabilísticas, baseadas nas leis de probabilidades, e as amostras não probabilísticas, que tentam reproduzir o mais fielmente possível a população alvo. Entretanto, somente as amostras probabilísticas podem, por definição, originar uma generalização estatística, apoiada no cálculo de probabilidades e permitir a utilização da potente ferramenta que é a inferência estatística; *n* tamanho da amostra

Estatística: Ciência da contagem; ciência que tem por objecto o agrupamento metódico dos fatos sociais que se prestam a uma avaliação numérica (população, natalidade, mortalidade, rendimento de impostos, produções agrícolas, criminalidade, religião, etc.); ramo das matemáticas aplicadas que recorre ao cálculo das probabilidades para estabelecer hipóteses com base em acontecimentos reais, com o fim de fazer previsões.

Média amostral: É uma variável aleatória, função dos valores da amostra, é definida como a soma de todos os valores da amostra dividido pelo número de observações da amostra. Serve para estimar a média populacional.

Média populacional: É o valor que representa um conjunto de valores da população. Definida como a soma de todos os valores da população dividido pelo número de observações. Por exemplo: renda *per capita* de um país, esperança de vida, renda familiar média, pontuação média na escala de atitudes em relação à Estatística, etc.

Parâmetro: É uma medida usada para descrever, de forma resumida, uma característica da população, Por exemplo, a média populacional (m), a proporção populacional (p), a variância populacional (s), o coeficiente de correlação (r), etc. Os parâmetros, via de regra, são valores desconhecidos e desejamos estimar, ou testar, a partir dos dados de uma amostra.

População: É um conjunto de elemento com características semelhantes.

Variância amostral (s^2): Serve para estimar a variância populacional.

Variável: É uma característica da população. Toda questão de pesquisa define um número de construções teóricas que o pesquisador quer associar. O grau de operacionalização destas construções não faz parte de um consenso. Por essa razão, a seção que trata das definições das variáveis deve permitir ao leitor avaliar a adequação dos instrumentos utilizados, as variáveis escolhidas e as construções teóricas descritas no quadro conceitual

Referências

COSTA,S.F. **Introdução ilustrada à estatística**.. 3ª edição. São Paulo: Habra, 1992

CAZORLA, I.M & SANTANA,E.R.S. **Tratamento da Informação para o ensino Fundamental e Médio.**. Bahia: Via Litterarum, 2006

DOWNING, Douglas. **Estatística Aplicada**. São Paulo: Saraiva, 2000.

FONSECA, Jairo Simon da. **Curso de Estatística** . São Paulo: Atlas, 1996

FREUND, Jonh E. **Estatística Aplicada: Economia, Administração e Contabilidade** . Porto Alegre: Bookman, 2000.

MAGALHÃES, Marcos Nascimento. **Noções de Probabilidade e Estatística** . São Paulo: editora da Universidade de São Paulo, 2005.

MEYER, Paul L. **Probabilidade Aplicações à Estatística**. Ao livro Técnico S.A, 1974

SILVA, Ermes Medeiros da. Et all. **Estatística para o curso de Administração e Ciências Contábeis** . São Paulo: Atlas, 1997.

SPIGEL, M.R. **Estatística**.São Paulo: Makron Books. 3ª edição, 1993

STEVENSON, William.J. **Estatística Aplicada à Administração**. São Paulo: Habra, 1996