

# ESTATÍSTICA USANDO EXCEL

## UNIDADE I

### REVISANDO O EXCEL

#### REVISANDO FUNÇÕES

- **Função Soma**

**Ex: =SOMA(A1:A8)**

A fórmula irá somar todos os valores que se encontram no endereço A1 até o endereço A8. Os dois pontos indicam até, ou seja, some de A1 até A8. A fórmula será sempre a mesma, só mudará os devidos endereços dos valores que você deseja somar.

Veja o outro exemplo:

	A	B	C	D	E
1	10	25	15	10	=SOMA(A1:D1)
2					
3					

Neste exemplo estamos somando todos os valores do endereço **A1** até o endereço **D1**. A fórmula seria digitada como no exemplo, e ao teclar **ENTER** o valor apareceria. No caso a resposta seria **60**.

- **Função Média**

Calcula-se a média de uma faixa de valores, após somados os valores e divididos pela quantidade dos mesmos.

Exemplo: Suponhamos que desejasse saber qual a média de idade numa tabela de dados abaixo:

	A	B	C
1	IDADE		
2	15		
3	16		
4	25		
5	30		
6	MEDIA IDADE	=MEDIA(A2:A5)	

- **Função Cont.Valores**

Conta-se o número de indivíduos numa determinada faixa de valores.

Exemplo: Suponhamos que desejasse saber o total de alunos desta turma que está listada na tabela abaixo: A resposta seria **4**.

	A	B	C
1	ALUNOS		
2	JOÃO		
3	FRANCISCO		
4	CARLOS		
5	MARIA		
6	TOTAL	=CONT.VALORES(A2:A5)	

- **Função Condicional SE**

Suponhamos que desejasse criar um Controle de Notas de Aluno, onde ao se calcular a média, ele automaticamente especificasse se o aluno fora aprovado ou não. Então Veja o exemplo abaixo. Primeiramente, você precisa entender o que desejar fazer. Por exemplo: quero que no campo situação ele escreva **Aprovado somente se o aluno tirar uma nota Maior ou igual a 7 na média**, caso contrário ele deverá escrever **Reprovado, já que o aluno não atingiu a condição para passar**. Veja como você deve escrever a fórmula utilizando a função do SE>

	A	B	C
1	ALUNO	MÉDIA	SITUAÇÃO
2	Márcio	7	=SE(B2>=7;"Aprovado";"Reprovado")
3			

**“Bem, agora que você lembrou como usar algumas funções básicas do Excel daremos início as nossas planilhas eletrônicas.”**

## EXERCÍCIO

Usando a planilha que recebeu do instrutor. Calcule e formate a planilha deixando-a como no exemplo abaixo.

	A	B	C	D	E	F	G
1	<b>FOLHA DE PAGAMENTO DE FUNCIONÁRIOS</b>						
2							
3	<b>NOME</b>	<b>CARGO</b>	<b>SAL. BASE</b>	<b>H. EXTRA</b>	<b>TOT. PROVENTOS</b>	<b>INSS</b>	<b>TOT. LIQUIDO</b>
4	CLAYTON	Engenheiro	5.900,00	1.100,00	7.000,00	350,00	6.650,00
5	JOÃO	Engenheiro	5.000,00	1.000,00	6.000,00	350,00	5.650,00
6	MARIA	Engenheiro	1.500,00	90,00	1.590,00	143,10	1.446,90
7	ANTONIO	Engenheiro	6.400,00	110,00	6.510,00	350,00	6.160,00
8	MATEUS	Engenheiro	3.500,00	700,00	4.200,00	350,00	3.850,00
9	VALDIR	Supervisor	1.900,00	100,00	2.000,00	220,00	1.780,00
10	AMADEUS	Supervisor	1.100,00	80,00	1.180,00	106,20	1.073,80
11	CLAUDIO	Supervisor	1.100,00	80,00	1.180,00	106,20	1.073,80
12	BERGNAN	Vigia	460,00	65,00	525,00	42,00	483,00
13	BENEDITO	Vigia	460,00	50,00	510,00	40,80	469,20
14	GERTUDES	Vigia	460,00	75,00	535,00	42,80	492,20
15	DANIELMA	Vigia	460,00	90,00	550,00	44,00	506,00
16							
17	<b>RESUMO DA FOLHA:</b>						
18							
19	Total de Funcionários	12					
20	Maior Salário	6.650,00					
21	Menor Salário	469,20					
22	Salário Médio	2.469,58					
23							
24	Tot. Sal. Engenheiros	22300					
25	Tot. Sal. Supervisores	4100					
26	Tot. Sal Vigia	1840					
27							
28	Hora Extra	3.540,00					
29	INSS	2.145,10					
30	Salário Base	28.240,00					
31	Total Bruto	31.780,00					
32	Total Líquido	29.634,90					

Meta do Mês: R\$ 30.000,00  
Resultado: Meta Atingida



## FÓRMULAS DA PLANILHA

- **Na Folha**

**Total de proventos** → Salário Base + Hora Extra

**\*INSS** → Fórmula condicional do SE, que será detalhada a seguir

**Total Líquido** → Total dos Proventos – INSS

- **No Resumo**

**Total de Funcionários** → Função CONT.VALORES

**Maior Salário** → Função MÁXIMO

**Menor Salário** → Função MÍNIMO

**Salário Médio** → Função MÉDIA

**Total Sal. Engenh.** → Função SOMASE

**Total Sal. Supervi.** → Função SOMASE

**Total Sal. Vigias.** → Função SOMASE

**Outras** → Função SOMA

- **Meta do Mês**

**Resultado** → Função do SE + Formatação condicional

- **Informações para o cálculo do INSS na Folha de Pagamento**

**\*INSS**

**Fórmula condicional do SE com as seguintes condições:**

1ª Condição → Se o total dos proventos for até 965.67, o desconto será de 8 %.

2ª Condição → Se o total dos proventos for de 965.68 até 1609.45, o desconto será de 9%.

3ª Condição → Se o total dos proventos for 1609.46 até 3218.90, o desconto será de 11%.

4ª Condição → Se o total dos proventos for acima de 3218.90 será descontado o teto de 350.00.

## GERENCIANDO UMA BASE DE DADOS COM UMA TABELA DINÂMICA

	A	B	C	D	E	F	G	H	I
1	ALUNOS DE ESTATÍSTICA DA UFRN - ANO 2007								
2	INFORMAÇÕES PESSOAIS								
3	Sexo	Idade	Lazer	Altura	Peso	Altura do pai	Altura da mãe	Peso do pai	Peso da mãe
4	F	18	Ativ. Culturais	156	49	165	163	65	59
5	F	25	Praia	160	50	170	150	70	50
6	F	37	Balada	165	67	180	170	90	60
7	F	26	Praia	149	41	169	158	73	59
8	F	21	Praia	170	61	160	162	73	68
9	F	19	Praia	165	65	183	162	90	75
10	F	20	Outros	156	49	166	160	71	64
11	F	22	Praia	173	55	168	150	86	77
12	F	20	Praia	162	52	171	156	89	80
13	F	22	Praia	162	53	169	155	80	59
14	M	21	Balada	174	79	170	160	72	64
15	M	18	Ativ. Esportivas	172	75	170	165	70	60
16	M	25	Balada	175	60	175	161	80	57
17	M	20	Balada	176	65	176	152	95	48
18	M	27	Outros	175	69	170	168	81	78
19	M	21	Ativ. Esportivas	174	68	172	168	98	70
20	M	40	Outros	163	61	171	160	90	62
21	M	24	Praia	176	83	160	158	80	61
22	M	22	Balada	181	73	177	160	77	50
23	M	22	Balada	175	74	165	160	62	60
24	M	25	Ativ. Esportivas	166	60	160	156	63	60
25	M	23	Balada	170	63	167	160	63	60
26	F	30	Praia	165	69	176	150	90	55
27	M	52	Ativ. Culturais	170	71	168	160	85	70
28	M	20	Ativ. Esportivas	171	72	170	152	84	62

Com o uso da Tabela Dinâmica podemos visualizar rapidamente a média de idade, a média de peso e a média de altura, podendo ser visualizado por sexo ou de forma geral. Visualizamos também quanto que o Lazer influencia nessas médias. Tudo isso, dentre outras possibilidades.

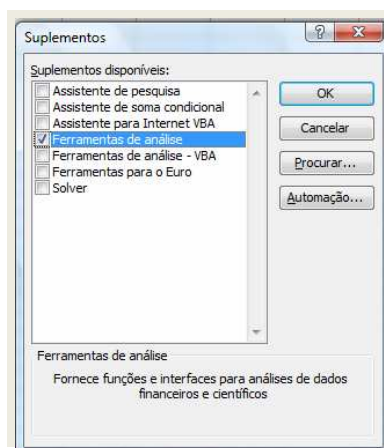
	A	B	C	D
1	Lazer	(Tudo)		
2				
3		Dados		
4	Sexo	Média de Idade	Média de Peso	Média de Altura
5	F	23	55	160
6	M	25	68	172
7	Total geral	24	63	167
8				

## UNIDADE II

### *Estatística Descritiva*

Algumas das *funções estatísticas* e todas as *ferramentas de análise* do Excel utilizadas neste curso, geralmente, não são incluídas ao iniciar o Excel. Sugerimos que o aluno realize a verificação seguinte antes de darmos início ao curso.

- No menu **Ferramentas** escolha **Suplementos**. O Excel apresentará a caixa de diálogo **Suplementos** com os **Suplementos disponíveis**.
- O suplemento **Ferramentas de análise** deve estar selecionado como mostra a figura seguinte.



O objetivo da *Estatística Descritiva* é organizar, resumir, analisar e interpretar observações disponíveis.

### 1 - VARIÁVEIS

*Variável* é uma característica da *unidade elementar* que pode ter valores diferentes entre as unidades medidas.

#### **Exemplo:**

A tabela seguinte registra parte do Cadastro de Funcionários da empresa. Quais são as unidades elementares e as variáveis deste cadastro?

Nome	Idade	Cargo	Sexo	Peso	Escolaridade
João	27	Supervisor	M	62 kg	2º Grau
Alex	38	Chefe	M	78 kg	3º Grau
Estela	34	Gerente	F	65 kg	3º Grau
Ana	32	Secretária	F	58 kg	1º Grau

**Solução.** As seis variáveis de cada funcionário da empresa compõem uma unidade elementar: Nome, Idade, Cargo, Sexo, Peso e Escolaridade.

## 1.1 - Classificação das variáveis

1.1.1 - Variáveis Quantitativas: Refere-se a quantidades medidas numa escala numérica.

As variáveis quantitativas podem ser do tipo: Discretas e Contínuas

- **Variáveis Discretas** – Referem-se a variáveis numéricas que assumem somente números inteiros positivos. **Ex:** *Quant. De vendas diárias de uma empresa; Quant. De peças defeituosas num lote de produção.*
- **Variáveis Contínuas** – Assumem qualquer valor dos números reais. **Ex:** *Valor da venda diária de uma empresa; Consumo mensal de energia elétrica.*

1.1.2 – Variáveis Qualitativas: Referem-se a variáveis não numéricas. As variáveis qualitativas poder ser classificadas como nominais e ordinais.

- **Variáveis nominais** – Não tem ordenamento nem hierarquia. **Ex:** *Sexo, Estado Civil.*
- **Variáveis ordinais** – São equivalentes as variáveis nominais, porém incluindo uma ordem. **Ex:** *Bom, Regular, Ruim; Primeira, Segunda, Terceira.*

1.1.3 – Seqüência Temporal. **Ex:** *As cotações diárias de uma ação na bolsa; A demanda de energia elétrica diária.*

## 2 – POPULAÇÃO E AMOSTRA

- O grupo completo de unidades elementares de pessoas, objetos ou coisas é denominado *população*.
- Um subconjunto de unidades elementares selecionados numa população é denominado *amostra*.

### 2.1 – Tipos de Pesquisa

- CENSO – Pesquisa que coleta a totalidade de uma população em busca da observação de determinadas características.
- AMOSTRAGEM – Pesquisa que coleta um pequeno grupo de indivíduos de uma população em busca da observação de determinadas características.

### Amostra

Muitas aplicações de estatística operam com amostras retiradas de uma população sobre a qual se deseja obter respostas. Simplesmente amostrar não é suficiente, a amostra deve ser *representativa* da população.

Uma *amostra representativa* tem as mesmas características da população de onde foi retirada.

## Gerando Dígitos Aleatórios com o Excel

O Excel 2003 dispõe das funções ALEATÓRIO e ALEATÓRIOENTRE para gerar *dígitos aleatórios* e da ferramenta de análise *Amostragem* para extrair amostras com reposição de uma população.

### ALEATÓRIO()

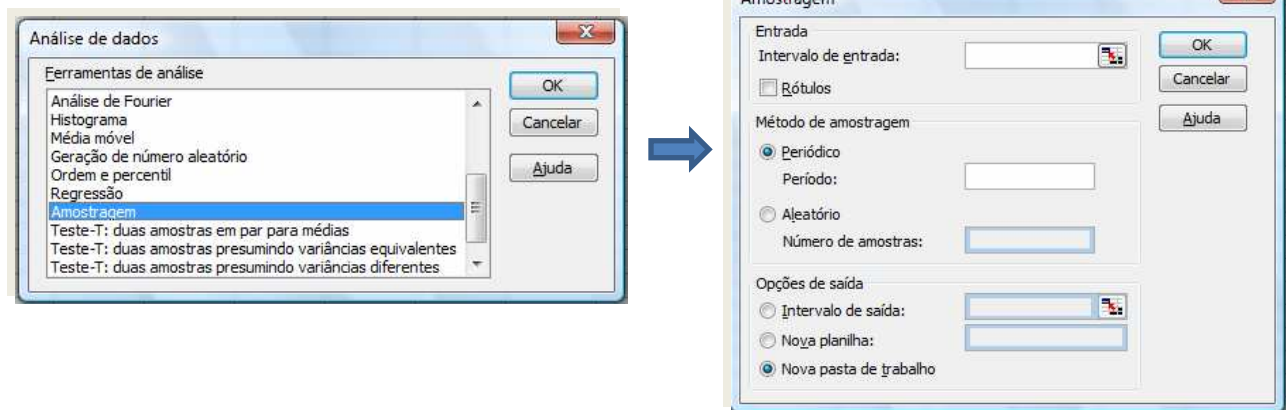
A função matemática ALEATÓRIO dá um grupo de dígitos aleatórios entre 0,00...0 e 1,00...0 com a quantidade de dígitos depois da vírgula definida pelo leitor, por exemplo: 0,236; 0,86945, etc.

### ALEATÓRIOENTRE(*inferior; superior*)

A função estatística ALEATÓRIOENTRE dá um grupo de dígitos aleatórios entre os grupos inferior e superior definidos nos argumentos da função.

Inserindo numa célula da planilha Excel a fórmula =ALEATÓRIOENTRE(0;599), o Excel fornecerá como resultado um grupo de dígitos entre o grupo 000 e o grupo 599.

## Ferramenta de Análise Amostragem



## 3 - TABELAS DE FREQUÊNCIA

A frequência do valor de uma variável é o número de repetições desse valor.

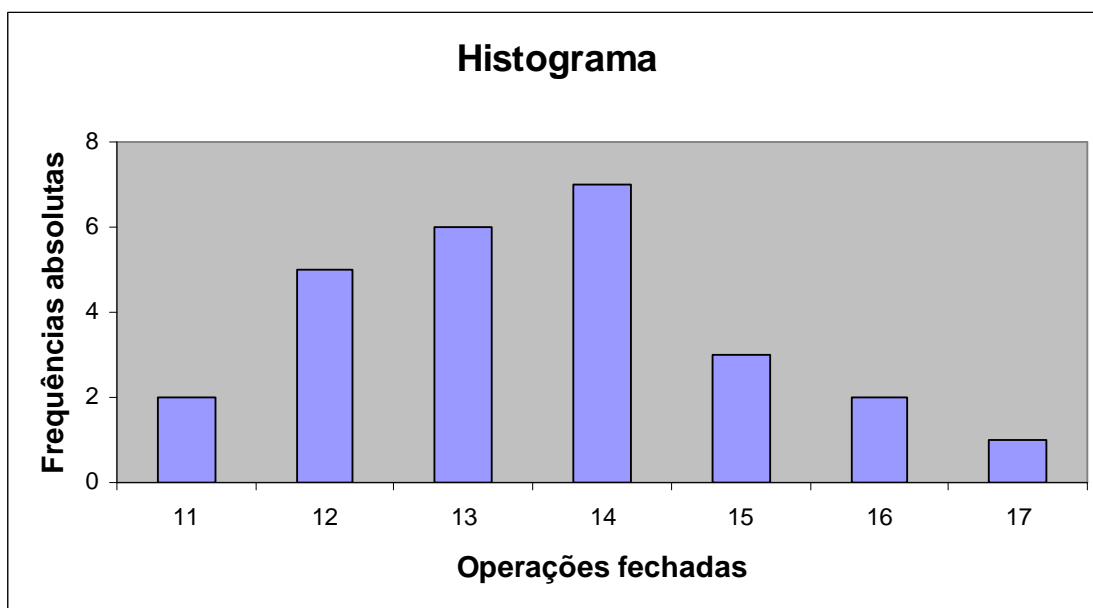
- Frequência Absoluta: É a contagem das repetições de cada valor dentro da variável;
- Frequência Relativa: É o resultado de dividir sua frequência absoluta pelo tamanho da amostra.
- Frequência Acumulada: É a soma das frequências absolutas ou relativas desde o valor inicial da variável.

Operações Fechadas por Dia	Frequências Absolutas	Frequências Relativas %	Frequência Acum. Absolutas	Frequência Acum. Relativa
11	2	7,69%	2	7,69%
12	5	19,23%	7	26,92
13	6	23,08%	13	50,00
14	7	26,92%	20	76,92
15	3	11,54%	23	88,46
16	2	7,69%	25	96,16
17	1	3,85%	26	100,00
<b>Total</b>	<b>26</b>	<b>100,00%</b>		

Tabela 1.1

## HISTOGRAMA

O *histograma* é uma forma gráfica de apresentar a distribuição de frequências de uma variável. O *histograma* é um gráfico de barras verticais construído com os resultados da tabela frequências.

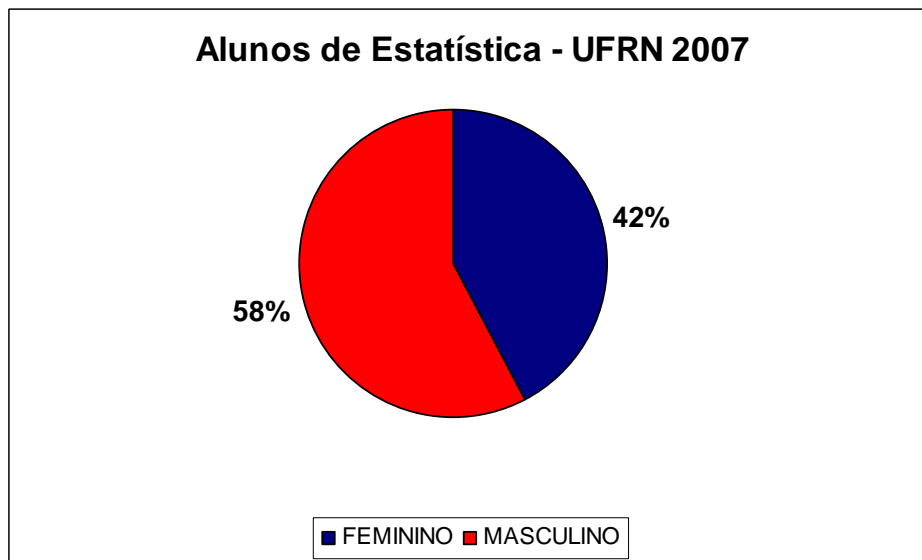


Fonte: Tabela 1.1

## GRÁFICO DE PIZZA

O *gráfico de pizza* deve ser usado preferencialmente para apresentar variáveis com no máximo 4 eventos.





Fonte: Base de Dados Alunos de Estatística – UFRN 2007.

### 3.1 - Distribuição de Frequências Utilizando Classes

- Fica complicado fazer tabela de frequência de variáveis contínuas, principalmente quando o número de elementos é muito grande.
- Não há regra exata para determinarmos o número de classes para análise de uma variável. Apenas orientações práticas para o analista.
- Serão necessárias algumas tentativas, como também, estamos sujeitos a cometer erros.
- Os valores da variável são transformados numa nova variável cujos mesmos estão compreendidos entre os limites dos intervalos das classes.

#### 3.1.1 - Cálculo para o tamanho dos intervalos

**N** = Tamanho da amostra

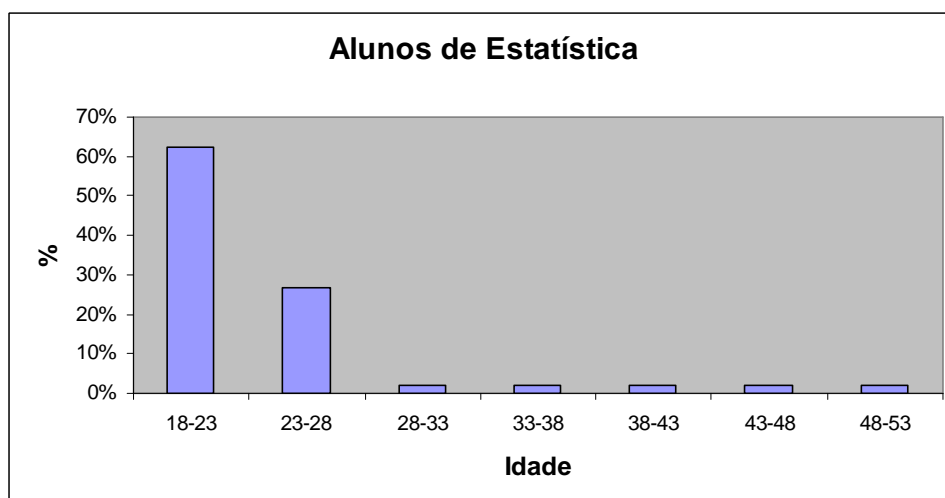
**K** (Número de classes)

$K = \sqrt{n}$       Obs. Arredonda para o mais próximo

**Intervalo de Variação** (Diferença do Max pelo Min)

$IV = \text{Max} - \text{Min}$

**Amplitude** =  $IV/K$



Fonte: Base de Dados Alunos de Estatística – UFRN 2007.

#### 4 – GRÁFICO DE PARETO

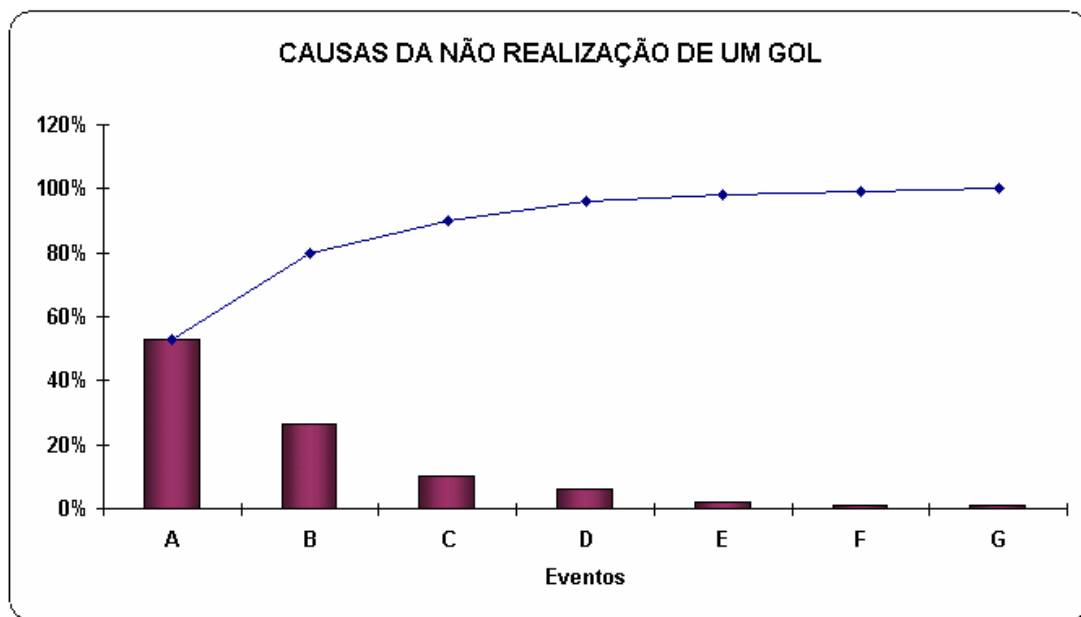
##### *Para que serve o Gráfico de Pareto?*

O gráfico de Pareto deve ser utilizado sempre que temos que estabelecer prioridades a partir de um número variado de informações e dados. Ajuda a dirigir nossa atenção e esforços para problemas verdadeiramente importantes, aumentando nossas chances de obtenção de bons resultados.

*Utilizaremos a tabela abaixo para demonstrar o gráfico de pareto*

Causas de não realização de um gol		
Eventos		Frequência
A	Passe errado	52
B	Chute errado a gol	26
C	Falta sofrida	10
D	Impedimento	06
E	Erro do juiz	02
F	Queda do Jogador	01
G	Defesa do goleiro	01
Total		98

Tabela 1.2



Fonte: Tabela 1.2

Assim, considerando o processo de melhoria da qualidade, o gráfico de Pareto é utilizado nas seguintes situações:

- Na seleção de problemas que serão como projetos de melhoria de qualidade;
- Na identificação das causas fundamentais de problema;
- Na comparação entre o antes e o depois de uma ação corretiva.

## 5 – MEDIDAS DE ORDENAMENTO

Com as tabelas de frequências e os métodos gráficos organizamos e começamos a análise dos valores de uma variável. Agora iniciamos os *métodos numéricos* para resumir e analisar os valores de uma variável, seja uma amostra ou a própria população.

Em algumas situações o interesse está em conhecer a *posição* de um determinado valor em relação ao grupo de valores. Por exemplo, qual a posição de um retorno de 15% ao ano com relação às rentabilidades anuais das aplicações do mercado financeiro? Quantos retornos do mercado financeiro são maiores que 15 %.

### 5.1 - Percentil

Um percentil é uma medida da posição relativa de uma unidade observacional em relação a todas as outras. O  $p$ -ésimo percentil tem no mínimo  $p\%$  dos valores abaixo daquele ponto e no mínimo  $(100 - p)\%$  dos valores acima.

Por exemplo,

- se uma altura de 1,80m é o 90o. percentil de uma turma de estudantes, então 90% da turma tem alturas menores que 1,80m e 10% têm altura superior a 1,80m.

- se o peso de uma pessoa de 75kg é o 40o. percentil de um conjunto de empregados. então 40% dos empregados pesam menos que 75kg e 60% pesam mais.

No Excel a função estatística PERCENTIL dá o valor que divide a *matriz* em duas partes, uma menor que o argumento *k* e a outra maior que *k*. O argumento *k* é um valor entre 0 e 1 correspondendo respectivamente a 0% e 100% do número de valores da *matriz*.

Alguns percentis da variável idade da Base de Dados Alunos Estatística - UFRN 2007.

Formula: =PERCENTIL(*Matriz*; *k*)

Percentil 10 = **19**

Percentil 20 = **20**

Percentil 50 = **22**

Percentil 82 = **26**

Percentil 94 = **38**

## 5.2 - Quartil

Os *percentis* dividem o intervalo de variação dos valores da variável em cem partes iguais. Dividindo os valores ordenados da variável em quatro quartos iguais se obtém três *quartis* denominados: *primeiro quartil*, *segundo quartil* e *terceiro quartil*. Numa variável com *n* valores ordenados de forma crescente:

- O primeiro quartil **Q1** é um valor tal que 25% dos valores da variável são menores e os restantes 75% são maiores.
- O segundo quartil **Q2** é um valor tal que 50% dos valores da variável são menores e os restantes 50% são maiores.
- O terceiro quartil **Q3** é um valor tal que 75% dos valores da variável são menores e os restantes 25% são maiores.

Quartis da variável idade da Base de Dados Alunos Estatística - UFRN 2007.

Formula: =QUARTIL(*Matriz*; *quarto*)

Quartil 1 (25%) = **20**

Quartil 2 (50%) = **22**

Quartil 3 (75%) = **22**

## 6 - MEDIDAS DE TENDÊNCIA CENTRAL

As medidas de tendência central nos dão uma noção sobre qual valor representa o centro de uma distribuição ordenada (rol), ou simplesmente um valor que melhor represente essa distribuição

## **6.1 Mediana**

A mediana é o valor localizado na posição central de uma distribuição de frequência ordenada. Sofre pouca influência pelas caudas da distribuição.

*Calculando a Mediana*

31 38 19 27 24 42 32 18 43 15 39
----------------------------------

*Ordena-se os valores*

15 18 19 24 27 <b>31</b> 32 38 39 42 43
---

*Mediana = 31*

Obs: Se a quantidade de valores n for um número par, a Md não será um valor da variável. A Md será o resultado da soma dos dois valores centrais dividido por 2.

<i>No Excel:</i> =MED(núm1;núm2;...;númN)
---

## **6.2 Moda**

A moda é o valor da variável que mais se repete. Sofre pouca influência pelas caudas da distribuição.

Moda é o valor da variável que mais se repete, ou seja, que tem maior frequência em uma variável.

15 38 42 27 24 42 32 18 43 15 42
----------------------------------

*Moda = 42*

<i>No Excel:</i> =MODA(núm1;núm2;...;númN)
--

## **6.3 Média**

Valor que representa uma distribuição. É a medida de posição mais usada. Sofre forte influência pelas caudas da distribuição.

Média da população é o resultado de dividir a soma dos valores  $X_1, X_2, \dots, X_N$ , da variável X pela quantidade de valores N.

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

<i>No Excel:</i> =MÉDIA(núm1;núm2;...;númN)
---

## 7 - MEDIDAS DE DISPERSÃO

Para identificar uma variável de forma completa deve-se adicionar uma medida numérica que mostre a variabilidade ou dispersão de seus valores. O conhecimento da dispersão dos valores é uma medida chave na análise estatística de uma variável.

### 7.1 – Variância

A *variância* de uma *variável aleatória* é uma medida da sua dispersão estatística, indicando quão longe em geral os seus valores se encontram do *valor esperado*.

Variância da população e variância da amostra

A variância de uma população  $y_i$  onde  $i = 1, 2, \dots, N$  é dada por

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (y_i - \mu)^2,$$

**No Excel:** =VARP(núm1;núm2;...;númN)

Quando estimando a variância da população usando  $n$  amostras aleatórias  $x_i$  onde  $i = 1, 2, \dots, n$ , a fórmula seguinte é um estimador não enviesado:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

**No Excel:** =VAR(núm1;núm2;...;númN)

### 7.2 – Desvio Padrão

O *Desvio Padrão* é a raiz quadrada da sua *variância*. Dessa maneira:

- O *desvio padrão* considerado como população é:  $\sigma_x = +\sqrt{\sigma_x^2}$

**No Excel:** =DESVPADP(núm1;núm2;...;númN)

- O *desvio padrão* considerado como amostra é:  $s_x = +\sqrt{s_x^2}$

**No Excel:** =DESVPAD(núm1;núm2;...;númN)

### **7.3 – Coeficiente de Variação**

O *coeficiente de variação* CV é o resultado de dividir o desvio padrão pela sua média. A medida relativa CV permite a comparação de distribuições, pois seu resultado é o desvio padrão por unidade de média. Comparando duas variáveis, a variável que tiver menor CV tem menor dispersão ou variabilidade.

## **8 – INCLINAÇÃO DE UMA DISTRIBUIÇÃO**

O *coeficiente de inclinação* é o resultado de comparar a distribuição de frequências dos valores de uma variável com uma distribuição normal.

<b>No Excel:</b> =DISTRORÇÃO(núm1;núm2;...;núm30)
---

O resultado da função DISTRORÇÃO deve ser interpretado da seguinte forma:

- Se o *coeficiente de inclinação* for igual a zero, a distribuição de frequências será simétrica.
- Se o *coeficiente de inclinação* for negativo, a distribuição de frequências terá inclinação esquerda ou negativa.
- Se o *coeficiente de inclinação* for positivo, a distribuição de frequências terá inclinação direita ou positiva.

## **9 – CURTOSE DE UMA DISTRIBUIÇÃO**

Se duas distribuições de frequências têm a mesma dispersão e inclinação não será suficiente para supor que as duas tenham a mesma forma, característica denominada curtose.

A curtose é medida pelo coeficiente de curtose que compara a distribuição de frequências de amostras com a distribuição normal.

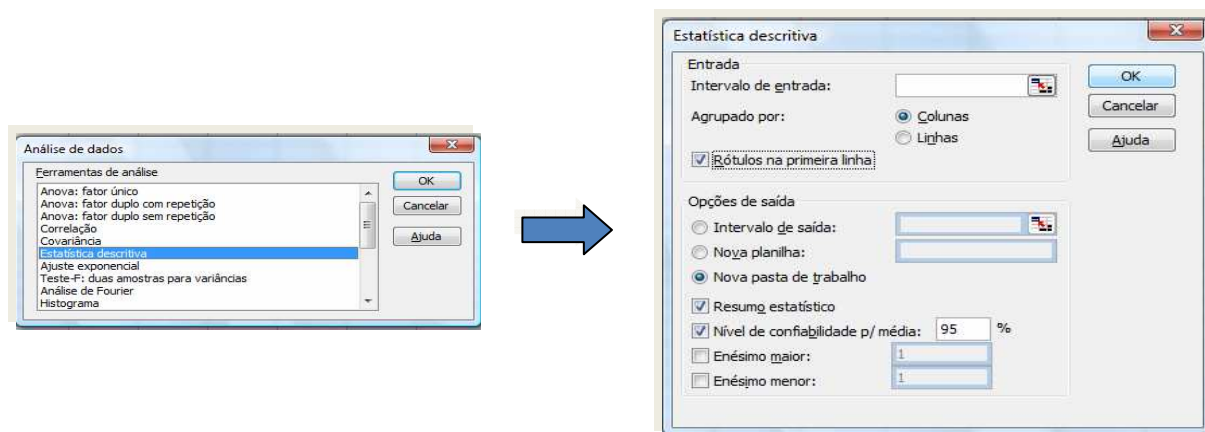
<b>No Excel:</b> = CURT(núm1;núm2;...;núm30)
--

O resultado desta função deve ser interpretado da seguinte forma:

- Se o *coeficiente de curtose* for igual a zero, então a distribuição de frequências será a própria distribuição normal.
- Se o *coeficiente de curtose* for negativo, então a distribuição será achatada, plana.
- Se o *coeficiente de curtose* for positivo, a distribuição será concentrada ao redor da média, distribuição com pico.

## 10 - FERRAMENTA DE ANÁLISE ESTATÍSTICA DESCRITIVA

No menu **Ferramentas** escolhemos **Análise de Dados** e recebemos a caixa de diálogo com todas as ferramentas de análise disponíveis. Depois de escolher *Estatística Descritiva* e pressionar o botão Ok o Excel apresenta a caixa de diálogo **Estatística descritiva**.



Para utilizar a ferramenta de análise *Estatística Descritiva* utilizaremos a **Variável Idade** da planilha Base de Dados Alunos de Estatística – UFRN 2007.

	A	B	C	D	E
1	<b>ALUNOS DE ESTATÍSTICA DA UFRN - ANO 2007</b>				
2					
3		Idade			
4		18			
5		18			
6		18			
7		19			
8		19			
9		19			
10		19			
11		19			
12		20			
13		20			
14		20			
15		20			
16		20			
17		20			
18		21			
19		21			
20		21			
21		21			
22		21			
23		21			
24		22			
25		22			

Idade		
Média		23,93333333
Erro padrão		1,060017153
Mediana		22
Modo		22
Desvio padrão		7,110811231
Variância da amostra		50,56363636
Curtose		7,072020722
Assimetria		2,598394671
Intervalo		34
Mínimo		18
Máximo		52
Soma		1077
Contagem		45
Nível de confiança(95,0%)		2,136324168



## 11 - BOXPLOT

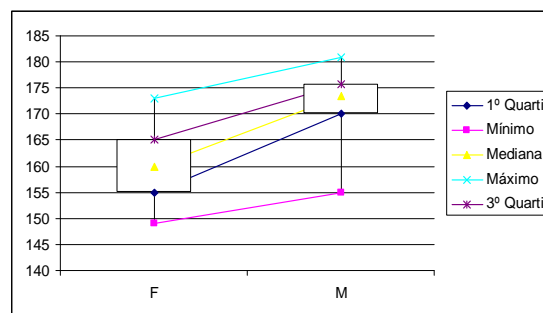
### 11.1 – Criando o boxplot no Excel

1. O primeiro passo é gerar as estatísticas para o conjunto de dados conforme descrito abaixo, e na mesma sequência.

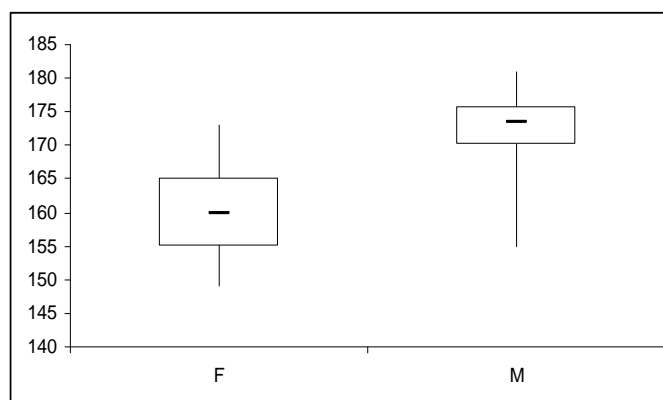
	Dados 1	Dados 2
1º Quartil	155	170
Mínimo	149	155
Mediana	160	173,5
Máximo	173	181
3º Quartil	165	175,75

2. Selecione toda a informação, incluindo os rótulos de dados e depois selecione **Inserir > Gráfico** e escolha **Linha**. Selecione o sub-tipo **‘linha com marcadores exibidos a cada valor de dado’**. Clique em **avançar**, e selecione **‘séries em: linhas’** e depois click em **‘concluir’**.

3. Agora serão executados comando sobre o gráfico. Clique com o botão direito do mouse sobre uma das linhas do gráfico. Selecione **‘formatar série de dados’**, vá para **‘opções’** e selecione **‘linhas de máximo/mínimo’** e **‘barras superiores/inferiores’**. A **‘largura do espaçamento’** poderá variar entre diversos valores, pois isso não importa para o Boxplot. O resultado deverá ser algo do tipo:

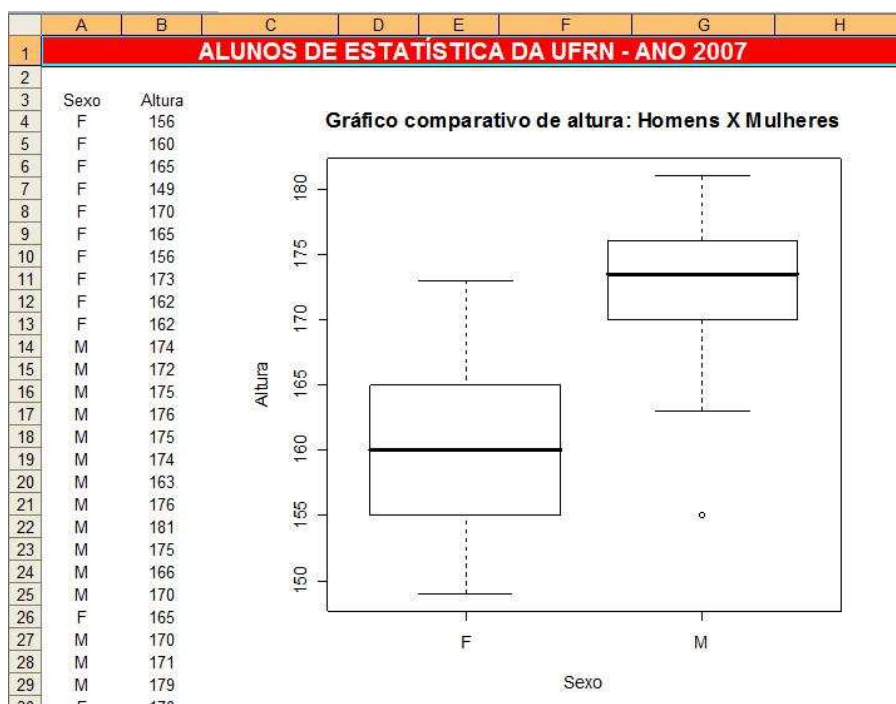


4. As linhas serão removidas uma a uma, inicialmente clicando com o botão direito do mouse sobre a linha. Selecionar **‘formatar série de dados’** em **‘padrões’** observar o campo **‘linha’** e marcar sobre **‘nenhuma’**, da mesma forma para o campo **‘marcador’**, deixando apenas o marcador para a **‘linha da mediana’**. Isso deve ser feito para todas as linhas. Em seguida, clique com o botão direito sobre a área do gráfico, selecione **‘opções de gráfico’**, na guia **‘linhas de grade’** desmarque todas as opções de linha e na **‘guia legenda’** desmarque a opção **‘mostrar legenda’**. O resultado final será como está abaixo.



## 11.2 Exercício:

Para utilizar o boxplot utilizaremos as variáveis **Sexo** e **Altura** da planilha Base de Dados Alunos de Estatística – UFRN 2007.



## 12 - ANÁLISE DE CORRELAÇÃO E REGRESSÃO LINEAR

- O Coeficiente de Correlação “R” é uma medida do grau de associação linear entre duas variáveis.
- O Coeficiente de Determinação “R<sup>2</sup>” é uma medida do poder explicativo do modelo utilizado.

### 12.1 – Resumo dos valores do Coeficiente de Correlação

A seguir apresentaremos o significado dos valores possíveis do coeficiente de correlação limitados no intervalo **-1 e +1**.

**r=+1** → *Correlação positiva perfeita*. Os pares de valores das duas variáveis estão numa mesma reta com declividade positiva. Neste caso, pode-se ver a perfeita relação linear entre as variáveis.

**r próximo de +1** → *Forte correlação positiva*. A maioria dos pares de valores está próxima de uma reta com declividade positiva.

**r próximo de +0** → *Fraca correlação positiva*. Os pares de valores formam uma nuvem com ligeira tendência de declividade positiva.

$r=0 \rightarrow$  Não há relação. Os pares de valores formam uma nuvem de pontos sem nenhuma tendência.

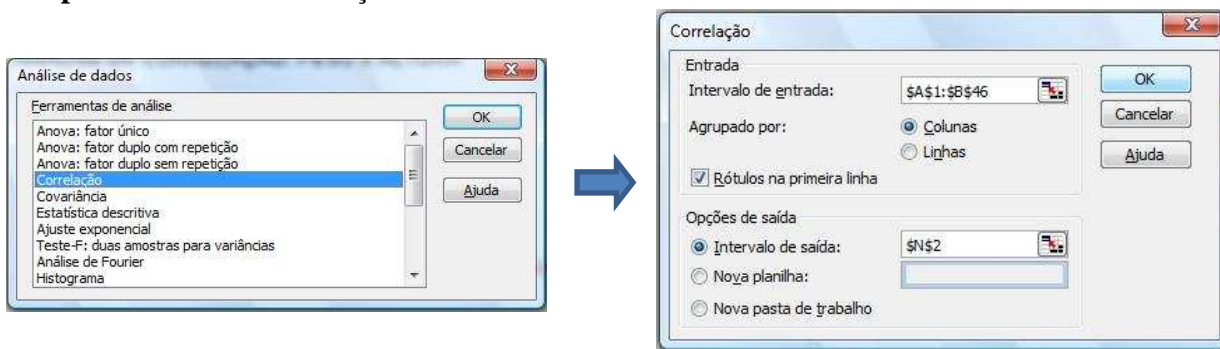
$r$  próximo de  $-0 \rightarrow$  Fraca correlação negativa. Os pares de valores formam uma nuvem com ligeira tendência de declividade negativa.

$r$  próximo de  $-1 \rightarrow$  Forte correlação negativa. A maioria dos pares está próxima de uma reta com declividade negativa.

$r=-1 \rightarrow$  Correlação negativa perfeita. Os pares de valores das duas variáveis estão numa mesma reta com declividade negativa. Neste caso, mais uma vez, pode-se ver a perfeita relação linear entre as observações.

No Excel: =CORREL(matriz1;matriz 2)

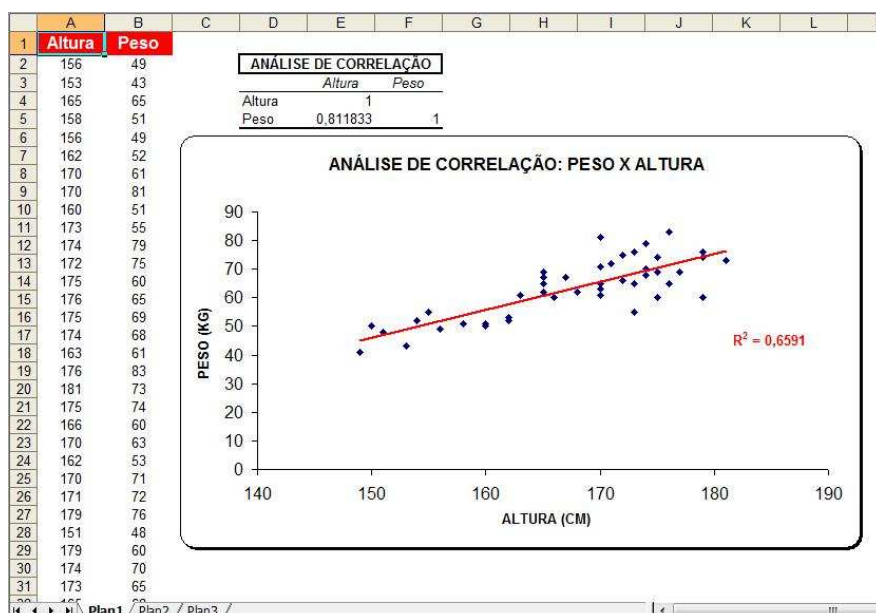
### Exemplo: Análise de correlação entre Peso e Altura



ANÁLISE DE CORRELAÇÃO		
	Altura	Peso
Altura	1	
Peso	0,811833	1

Exemplo com a base de dados Alunos de Estatística da UFRN – Ano 2007.

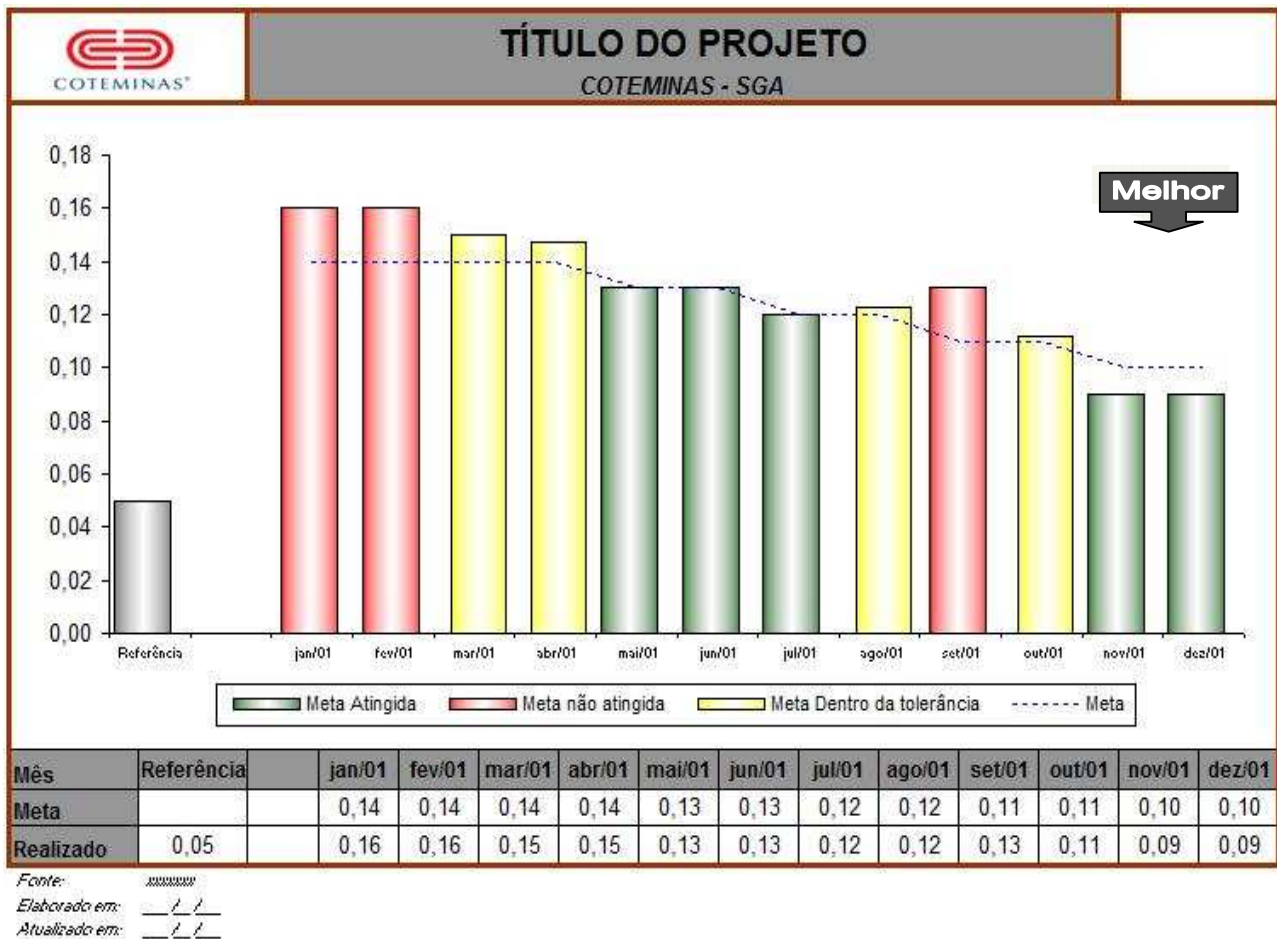
### Exibindo Gráfico de Dispersão com a Linha de Tendência



## UNIDADE III

### APLICAÇÕES

#### 1 – Criando um gráfico dinâmico para acompanhamento de processos.



## 2 – Transformando uma planilha comum em um Banco de Dados.

### Relatório de Anomalias - Planilha

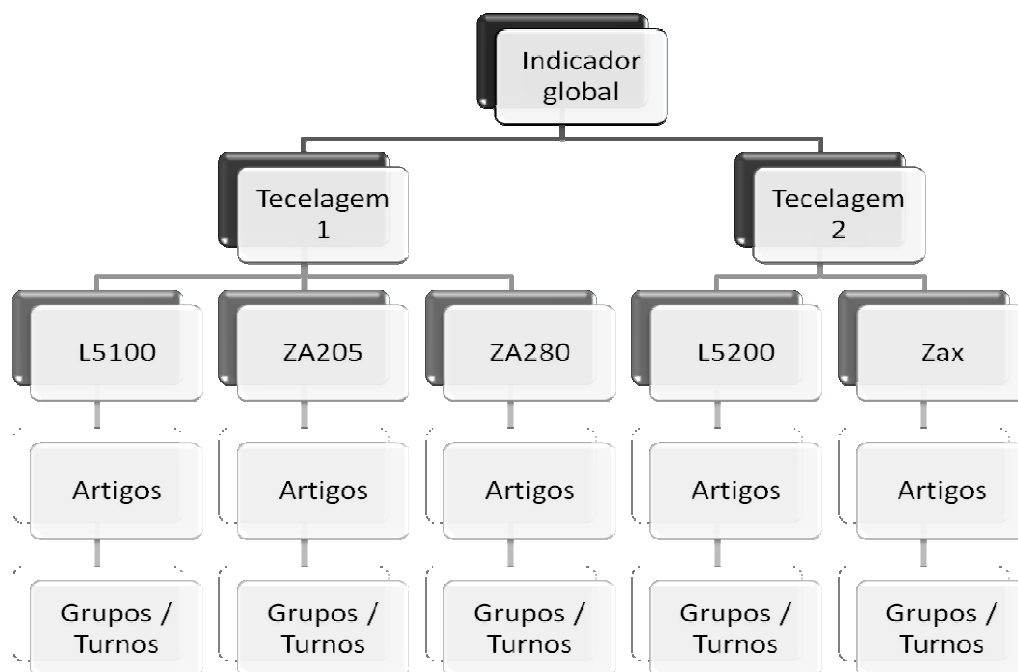
<b>Relatório de Anomalia</b>				Data:	
				Turno:	
Área:				Responsável:	
Equipamento:				Operador:	
<b>DESCRIÇÃO RESUMIDA DO FATO</b>					
DATA INICIO	HORA INICIO	DATA TERMINO	HORA TERMINO		
<b>OCORRÊNCIA</b>					
<input type="checkbox"/> Defeito Mecânico <input type="checkbox"/> Falha Operacional <input type="checkbox"/> Defeito Elétrico <input type="checkbox"/> _____					
<b>ANÁLISE DE CAUSAS</b>					
CAUSA					
POR QUE?					
POR QUE?					
POR QUE?					
POR QUE?					
POR QUE?					
CAUSA RAIZ					
<b>CONTRAMEDIDAS</b>					
Ação	Responsável	Prazo	Observações		

### Relatório de Anomalias – Banco de Dados

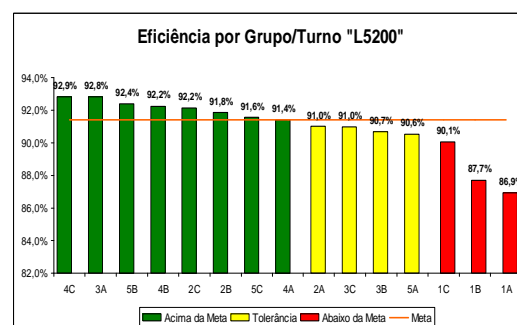
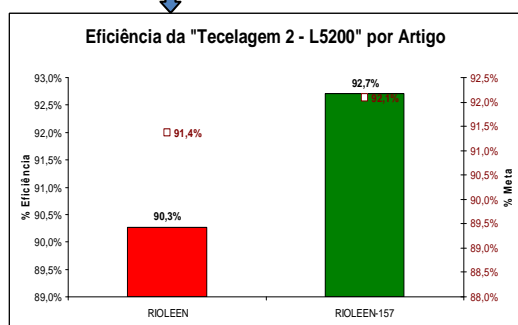
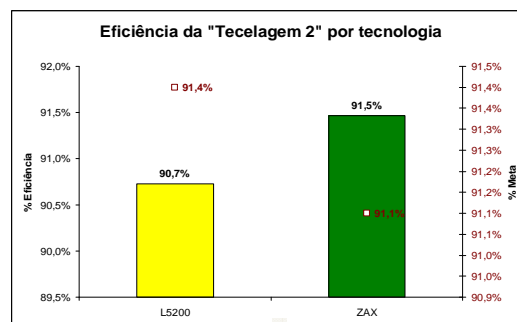
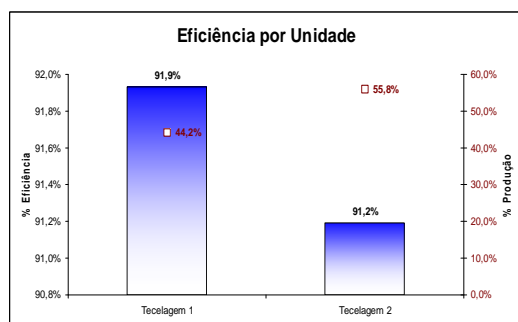
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	RELATÓRIO DE ANOMALIAS															
2	Data	Mês	Sem	Ter	Ar	Responsável	Equipamento	Operador	Fato	Data/Hora Início	Data/Hora Término	Dura	Ocorrência	Causa Raiz	Ação	Prazo / Responsável
3	13/07/2009	7	Semana 2	A	XXX	JOAO	MAQUINA DE LAVAGEM	MAHOEL	VAZANDO ÁGUA	13/07/2009 10:20	13/07/2009 13:30	3:10:00	DEFEITO MECÂNICO	MATERIAL VENCIDO	TROCA DO MATERIAL	15/07/2009
4	13/07/2009	7	Semana 2	A	YYY	JOSE	ESTEIRA	ROMARIO	TRAVOU	13/07/2009 11:00	13/07/2009 15:25	4:25:00	DEFEITO MECÂNICO	FALTA DE LUBRIFICAÇÃO	LUBRIFICAÇÃO	14/07/2009
5	14/07/2009	7	Semana 2	B	XXX	JOAO	TREM DE POUSO	MAZINHO	TRAVOU	14/07/2009 15:00	14/07/2009 21:00	6:00:00	DEFEITO MECÂNICO	MANUTENÇÃO PREVENTIVA	IMPLANTAR PROGRAMA	24/07/2009
6	14/07/2009	7	Semana 2	C	YYY	JOSE	PAINEL DE CONTROLE	RAI	APAGOU	14/07/2009 23:50	15/07/2009 05:35	5:45:00	DEFEITO ELÉTRICO	FALHA NOS CABOS ELÉTRICO	TROCA DE CABOS ELET	17/07/2009
7	15/07/2009	7	Semana 2	C	XXX	JOAO	MAQUINA DE LAVAGEM	MAHOEL	SUJEIRA NOS TEC	16/07/2009 02:00	16/07/2009 06:00	4:00:00	DEFEITO MECÂNICO	FILTRO NÃO FUNCIONA	TROCA DO FILTRO	16/07/2009
8	15/07/2009	7	Semana 2	C	XXX	JOAO	COMPRESSOR	MARADONA	AQUECIMENTO DE	15/07/2009 22:55	16/07/2009 03:35	4:40:00	DEFEITO MECÂNICO	DUTOS DE ÁGUA FRIA	TROCA DOS DUTOS	20/07/2009
9	16/07/2009	7	Semana 2	A	YYY	JOSE	ESTEIRA	ROMARIO	TRABALHANDO	16/07/2009 07:25	16/07/2009 11:15	3:50:00	DEFEITO MECÂNICO	FALTA DE REPARO NO MOTOR	CORRIGIR PROGRAMA	22/07/2009
10	16/07/2009	7	Semana 2	B	YYY	JOSE	PAINEL DE CONTROLE	RAI	FALHA NA PROG	16/07/2009 15:00	16/07/2009 22:00	7:00:00	FALHA OPERACIONAL	FALTA DE TREINAMENTO DOS	IMPLANTAR PROGRAM	15/09/2009
11	16/07/2009	7	Semana 2	C	XXX	JOAO	EMPILHADEIRA	PELE	NÃO FUNCIONA	16/07/2009 23:45	17/07/2009 03:00	3:15:00	OUTROS	FALTA DE PEÇAS	COMPRA DE PEÇAS	25/07/2009
12	17/07/2009	7	Semana 2	B	XXX	JOAO	EMPILHADEIRA	PELE	NÃO FUNCIONA	17/07/2009 15:10	17/07/2009 20:00	4:50:00	OUTROS	FALTA DE PEÇAS	COMPRA DE PEÇAS	25/07/2009
13	17/07/2009	7	Semana 2	B	YYY	JOSE	COMPUTADOR CENTRAL	EDMUNDO	NÃO FUNCIONA	17/07/2009 16:20	17/07/2009 22:00	5:40:00	DEFEITO MECÂNICO	EQUIPAMENTO MUITO ANTIGO	ATUALIZAÇÃO DO EQUIP	30/07/2009
14	17/07/2009	7	Semana 2	C	YYY	JOSE	BOMBA DE ÁGUA	RONALDO	NÃO FUNCIONA	18/07/2009 01:00	18/07/2009 05:50	4:50:00	DEFEITO MECÂNICO	VÁLVULAS DESGASTADAS	SUBSTITUIÇÃO DAS VÁ	19/07/2009
15	18/07/2009	7	Semana 2	A	YYY	JOSE	ENGOMADEIRA	ROMARIO	PRINCIPIO DE INC	18/07/2009 08:00	18/07/2009 08:10	0:10:00	DEFEITO ELÉTRICO	FIÇÃO ANTIGA	TROCA DE CABOS ELET	20/07/2009
16	18/07/2009	7	Semana 2	A	XXX	JOAO	TREM DE POUSO	MAZINHO	TRAVOU	18/07/2009 10:00	18/07/2009 13:00	3:00:00	DEFEITO MECÂNICO	MANUTENÇÃO PREVENTIVA	REVER O PROGRAMA D	24/07/2009
17	18/07/2009	7	Semana 2	B	XXX	JOAO	PAINEL DE CONTROLE	ZAGALO	DESCONFIGURAC	18/07/2009 15:00	18/07/2009 19:35	4:35:00	DEFEITO ELÉTRICO	BATERIA DESCARREGADA	TROCA BATERIA	19/07/2009
18	18/07/2009	7	Semana 2	C	XXX	JOAO	ESTEIRA	CAFU	LENTA	18/07/2009 22:50	19/07/2009 05:00	6:10:00	DEFEITO ELÉTRICO	MOTOR SEM POTENCIA	TROCA DO MOTOR	20/07/2009
19																

### 3 - Montando uma Árvore de Análises para definir qual o caminho indicativo que deverá ser seguido na hora de detalhar os níveis da informação.

#### ÁRVORE DE ANÁLISES



#### Gráficos seguindo a ordem da Árvore



## REFERÊNCIAS BIBLIOGRÁFICAS

LAPPONI, Juan Carlos. **Estatística Usando Excel**. Quarta Edição. Elsevier. Editora Campus.

**E**statística Usando Excel ensina Estatística e explica como aplicar os conceitos e analisar resultados por meio de exemplos resolvidos com os procedimentos tradicionais de cálculo e o Excel. Essa integração torna o aprendizado de Estatística eficaz e mostra que há diversos caminhos para alcançar o mesmo resultado, incluindo a superposição de recursos do Excel.

*LAPPONI, Juan Carlos.*