



# ESTATÍSTICA DESCRITIVA E AMOSTRAGEM

**Prof. Dr. Ednaldo Carvalho Guimarães**

E-mail: [ecg@ufu.br](mailto:ecg@ufu.br)  
tel: 3239-4156 ou 9102-9990

Curso de Pós-Graduação  
Especialização em Estatística Empresarial

---

---

---

---

---

---

---

---



## PROGRAMA

- Conceitos básicos
- Análise gráfica e construção de tabelas (distribuição de frequências)
- Medidas de posição
- Medidas de dispersão
- Análise de valores discrepantes
- Assimetria e Curtose
- Técnicas de Amostragem

---

---

---

---

---

---

---

---



## CRITÉRIOS DE AVALIAÇÃO

- Exercícios avaliativos individuais e/ou em grupo que serão resolvidos em sala de aula ou resolvidos extra classe, mas com data definida de entrega.

---

---

---

---

---

---

---

---



## BIBLIOGRAFIA

- BOLFARINE H. e BUSSAB W. O. *Elementos de Amostragem*. São Paulo: Ed. Blücher, 2005.
- BUSSAB, W. O.; MORETTIN, P. *Estatística Básica*. São Paulo : Atual, 2002.
- COSTA NETO, P. L. de O. *Estatística*. São Paulo : Edgard Blücher, 2002.
- FREUD, J. E.; SIMON, G. A. *Estatística aplicada*. Bookman, 2000, 403 p..
- KAZMIER, L. J. *Estatística Aplicada à Administração e Economia*, Rio de Janeiro: Makron Books, 1982.
- LEVINE, D. M.; BERENSON, M. L.; STEPHAN, D. *Estatística: teoria e aplicações (usando o Microsoft Excel em português)*. LTC editora, 2000, 812 p.
- SPIEGEL, M. R. *Estatística*. 3ª Ed. São Paulo, Markon Books , 1993. 642 p.
- STENVENSON, W. J. *Estatística Aplicada à Administração*. Harbra, 1986.
- TRIOLA, M. F. *Introdução à estatística*. Rio de Janeiro: LTC. 7ª edição, 1999.

---

---

---

---

---

---

---

---

---

---



## ESTATÍSTICA

- **Vem do latim “status” = Estado**
- **inicialmente envolvia:**
  - ✓ compilações de dados e gráficos representativos dos vários aspectos de um estado ou país.
    - taxa de mortalidade,
    - taxa nascimento,
    - renda,
    - taxas de desemprego, etc.

---

---

---

---

---

---

---

---

---

---



## MÉTODOS CIENTÍFICOS

**MÉTODO** é um conjunto de meios dispostos convenientemente para se chegar a um fim que se deseja.

- **Método Experimental:** consiste em manter constantes todas as causas (fatores), menos uma, e variar esta causa de modo que o pesquisador possa descobrir seus efeitos, caso existam.
- **Método Estatístico:** Admite todas essas causas presentes variando-as, registrando essas variações e procurando determinar, no resultado final, que influências cabem a cada uma delas.

---

---

---

---

---


---

---

---

---

---


 Universidade Federal de Uberlândia  
 Faculdade de Matemática  
 Curso de Especialização em Estatística Empresarial

## ESTATÍSTICA

É uma coleção de métodos para:

- ✓ planejar experimentos,
- ✓ obter dados,
- ✓ organizar,
- ✓ resumir,
- ✓ analisar
- ✓ concluir sobre as informações coletadas

---

---

---

---

---


---

---

---

---

---


 Universidade Federal de Uberlândia  
 Faculdade de Matemática  
 Curso de Especialização em Estatística Empresarial

## A ESTATÍSTICA

### ESTATÍSTICA DESCRITIVA

- Parte da Estatística que apenas coleta, descreve, organiza e analisa um conjunto de dados. Nela não são tiradas conclusões.

### ESTATÍSTICA INDUTIVA

- Também é chamada de *inferência estatística*. A partir da análise de dados são tiradas conclusões.

---

---

---

---

---


---

---

---

---

---


 Universidade Federal de Uberlândia  
 Faculdade de Matemática  
 Curso de Especialização em Estatística Empresarial

### Aplicação da Estatística

- Pode-se verificar registros de aplicações de métodos estatísticos desde a antiguidade, principalmente em dados relativos aos Estados (Impérios), como levantamentos populacionais, estimativas de produção, etc..
- No século XX a Ciência Estatística teve grande impulso.
- Atualmente exerce papel primordial em todas as áreas de conhecimento, passando pelas áreas de ciências exatas, ciências biológicas e ciências humanas.

---

---

---

---

---

---

---

---

---

---



## OBJETIVOS DE UMA ANÁLISE EXPLORATÓRIA

- Conhecer as variáveis
- Verificar o comportamento geral dos dados
- Verificar a ocorrência de valores atípicos
- Resumir as informações em valores característicos e representativos dos dados

*Obs: Muitas vezes a análise descritiva é relegada para um segundo plano, mas é essencial para conclusões coerentes sobre a variável em análise.*

---

---

---

---

---

---

---

---



## VARIÁVEIS ESTATÍSTICAS

- Quando estudamos uma determinada população, estamos interessados em uma ou algumas características dessa população. Estas características variam de acordo com o elemento ou indivíduo selecionado e, portanto, são chamadas de variáveis.
- Pode-se classificar as variáveis para o estudo estatísticos em **Variáveis Qualitativas e Variáveis Quantitativas**.

---

---

---

---

---

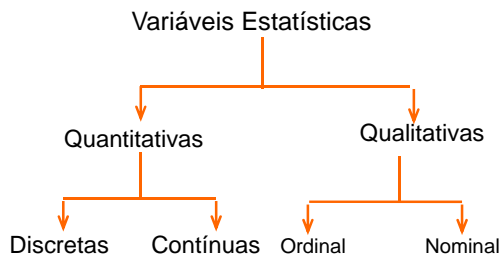
---

---

---



## VARIÁVEIS ESTATÍSTICAS



---

---

---

---

---

---

---

---



## ESCALA DE MEDIÇÃO

- **Escala Nominal.** Valores numéricos numa escala nominal apenas dão nome a uma categoria ou classe; os números são utilizados somente para diferenciar objetos, categorias ou nomes. Por exemplo, numa pesquisa de mercado realizada nas regiões sul e sudeste do Brasil, a variável *unidade da federação* do entrevistado foi codificada da seguinte forma: 1=Rio Grande do Sul, 2=Santa Catarina, 3=Paraná, 4=São Paulo, 5=Rio de Janeiro, 6=Minas Gerais e 7= Espírito Santo.

---

---

---

---

---

---

---

---



- **Escala Ordinal.** Valores numa escala ordinal dão nome e ordem a um objeto, categoria ou classe; os números são utilizados para diferenciar em ordem de superioridade, seguindo algum critério de hierarquia. Por exemplo, numa pesquisa, a variável *instrução do entrevistado* foi codificada assim: 1=Sem Instrução, 2=Primeiro Grau, 3=Segundo Grau, 4=Terceiro Grau, 5=Mestre e 6=Doutor.

---

---

---

---

---

---

---

---



- **Escala de Intervalos.** Valores numa escala de intervalos eliminam a limitação da escala ordinal estabelecendo intervalos iguais em que é possível ordenar as medições e, ao mesmo tempo, explicar quanto uma observação difere da outra. Por exemplo, o aumento da temperatura de ontem para hoje é de cinco graus, de 20 para 25 graus centígrados.

---

---

---

---

---

---

---

---



- **Escala Proporcional.** Valores numa escala proporcional eliminam a limitação da escala intervalar, estabelecendo um zero da própria categoria, denominado zero absoluto.

---

---

---

---

---

---

---

---



**Não adianta forçar os dados estatísticos pois o resultado pode ser inesperado**



---

---

---

---

---

---

---

---



### População e Amostra

▪ **QUAIS SÃO AS PREFERÊNCIAS MUSICAIS DOS ESTUDANTES BRASILEIROS ?**  
População - Amostra - Unidade estatística

- **População:** todos os estudantes brasileiros
- **Amostra:** grupo reduzido de estudantes
- **Unidade Estatística:** cada um dos estudantes

---

---

---

---

---

---

---

---



▪ **QUAIS SÃO AS PREFERÊNCIAS MUSICAIS DOS ESTUDANTES DA UFU?**

População - Amostra - Unidade estatística

- **População:** todos estudantes da UFU
- **Amostra:** grupo reduzido de estudantes da UFU
- **Unidade Estatística:** cada um dos estudantes

---

---

---

---

---

---

---

---



▪ **QUAL A RENDA DAS FAMÍLIAS DE UBERLÂNDIA?**

População - Amostra - Unidade estatística

- **População:** todas as residências de Uberlândia
- **Amostra:** grupo reduzido de residências
- **Unidade Estatística:** cada residência

---

---

---

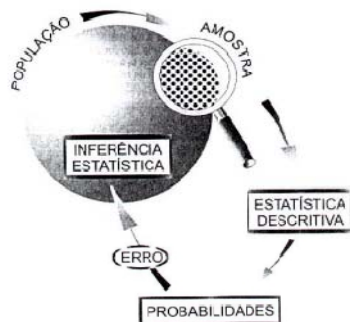
---

---

---

---

---



---

---

---

---

---

---

---

---



## Censo e Amostragem

### CENS

Num **C**enso são observados todos os indivíduos da população relativamente aos diferentes atributos que estão sendo objetos de estudo.

### AMOSTRAGE

Numa **M**ostragem, o estudo estatístico baseia-se numa parte da população, isto é, numa amostra que deve ser representativa dessa população (manter as características básicas da população)

---

---

---

---

---

---

---

---



### ▪ Por que o recurso de uma amostra e não de um censo ?

- Economia de tempo
- Redução de custos
- Impossibilidade de avaliar todos os elementos

### ▪ Como selecionar as amostras?

- Amostragem aleatória
- Amostragem sistemática
- Amostragem estratificada
- Amostragem por conglomerado
- Amostras não probabilísticas

---

---

---

---

---

---

---

---



## DISTRIBUIÇÃO DE FREQUÊNCIAS E ANÁLISES GRÁFICAS

- **Distribuição de frequências** → disposição das observações de uma variável em uma tabela
- **Análise gráfica** → representação da distribuição de frequências ou do comportamento de variáveis em figuras.

### Definições

- **Dados brutos** → sem ordenação
- **Dados ordenados ou rol** → ordem crescente ou decrescente de ocorrência

---

---

---

---

---

---

---

---



- **Amplitude Total** →  $A = R = \text{Max} - \text{Min}$
- **Frequência absoluta** → contagem do número de elementos da amostra ou da população que ocorre em determinada classe ou categoria → simbologia:  $f_i$  ou  $f_{ai}$  →  $f_i = n_i$ , sendo  $f_i$  a frequência da classe  $i$  e  $n_i$  o número de observações da classe  $i$ .
- **Frequência relativa** → número entre zero e um que é obtido dividindo-se a frequência absoluta de uma determinada classe pelo total de observações → simbologia:  $f_r$  → 
$$f_r = \frac{f_i}{\sum_i f_i}$$

---

---

---

---

---

---

---

---



**Frequência percentual** → é um número entre zero e 100. É a frequência relativa expressa em porcentagem → simbologia:  $f_p$  →  $f_{p_i} = 100 f_{r_i}$

*Obs: A vantagem de se utilizar a frequência percentual está na interpretação de resultados*

**Frequência acumulada** → comportamento da variável acima ou abaixo de um determinado valor ou de uma classe. Pode-se ter as frequências acumuladas absoluta, relativas ou percentuais, que serão denotadas, neste curso, respectivamente por:  $F$ ,  $F_r$  e  $F_p$ . Estas frequências são obtidas de modo semelhante às frequências simples.

---

---

---

---

---

---

---

---



## TABELAS E GRÁFICOS

**Observação:** Toda tabela e todo gráfico deve possuir títulos, sendo que o título da tabela deve estar acima da mesma e no caso dos gráficos, o título deve ser escrito abaixo dos mesmos.

### i) Variável qualitativa nominal ou ordinal

- A construção de tabelas consiste na definição das categorias e na determinação das respectivas frequências.
- As representações gráficas mais utilizadas neste caso são os gráficos do tipo barras ou gráficos de setores.

---

---

---

---

---

---

---

---





### ii) Variável quantitativa discreta

- A distribuição de frequências, quando não há uma amplitude muito grande na observações, é feita contando-se o número de ocorrências para cada observação.
- No caso de amplitude grande pode-se adotar os procedimentos que veremos para variáveis contínuas
- o tipo de gráfico mais apropriado para variáveis discreta é gráfico tipo barras.

---

---

---

---

---

---

---

---



**Exemplo :** Vamos construir uma tabela de frequências para os dados da variável número de pessoas residentes por domicílio, considerando uma amostra de 40 residências do conjunto Residencial Floresta.

4 4 4 5 4 1 2 3 6 4 6 4 4 6 3 5 3 4 4 4  
5 5 5 4 8 4 5 3 4 5 5 2 5 2 6 8 3 5 5 3

---

---

---

---

---

---

---

---



### DADOS

Tabela 2. Distribuição de frequências do número de pessoas residentes por domicílio, numa amostra de 40 residências do Conjunto Residencial Floresta, Uberlândia

Nº de pessoas	Frequência de residências	Frequência acumulada	Frequência percentual
1	1	1	2,5%
2	3	4	7,5%
3	6	10	15%
4	13	23	32,5%
5	11	34	27,5%
6	4	38	10%
7	0	38	0%
8	2	40	5%
Total	40		100%

---

---

---

---

---

---

---

---

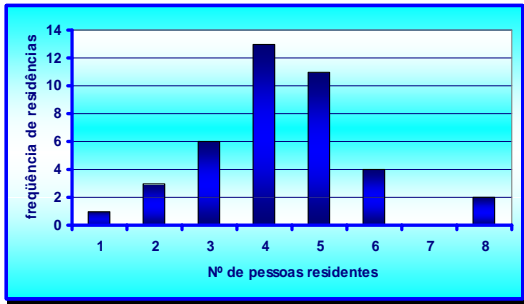


Figura 3. Gráfico de barras do número de pessoas por residência do condomínio Floresta.

---

---

---

---

---

---

---

---

### iii) Variável quantitativa contínua

- idéia de continuidade → criação de classes, ou seja, intervalos de ocorrências.
- as classes não necessariamente precisam ter a mesma amplitude
- a definição de variável contínua é feita em relação a variável e não em relação aos valores assumidos por ela em um levantamento de dados.
- Várias maneiras de se montar as classes
- recomenda-se que o número de classes esteja entre 5 e 20.

---

---

---

---

---

---

---

---

- É possível também ter distribuição de freqüências com diferentes intervalos de classes, isso ocorre quando existem classes com baixas freqüências em relação as demais e estas são unidas com outras classes.

### Simbologia das classes

Várias são as simbologias adotadas para as classes. Neste curso adotaremos a seguinte simbologia:

Seja A o limite inferior da classe i e B o limite superior da classe i, tem-se:

$A | \text{---} B \Rightarrow A \in \text{à classe e } B \notin \text{à classe.}$

---

---

---

---

---

---

---

---



**Sugestão de procedimento para a construção da distribuição de freqüências**

- a) Ordenar os dados
- b) Calcular a amplitude total:  $A = \text{Max} - \text{Min} \sqrt{n}$
- c) Determinar o número de classes (k):  $k = \sqrt{n}$  se  $n \leq 400$  ou  $k = 3 \cdot \log n$  se  $n > 400$ . Deve-se utilizar apenas a parte inteira de k.
- d) Determinar a amplitude de classe (c):  
$$c = A / (k - 1)$$

---

---

---

---

---

---

---

---



- e) Determinar o limite inferior da primeira classe ( $LI_{1a}$ ):  $LI_{1a} = \text{Min} - (c/2)$
- f) Determinar os demais limites de classes: O limite superior da primeira classe é dado por limite inferior da primeira classe mais a amplitude de classe; o limite inferior da segunda classe será igual ao limite superior da primeira; o limite superior da segunda será igual ao limite inferior da segunda mais a amplitude de classe e assim sucessivamente.
- g) Organizar a tabela e contar as ocorrências.

---

---

---

---

---

---

---

---



Tempo (seg) de início para o atendimento de uma central de relacionamento com clientes.

41	41	42	44	45	46	46	50
50	50	51	51	52	52	54	57
58	58	60	60				

---

---

---

---

---

---

---

---



### Construção de uma distribuição de frequências

- Organizar os dados
- Calcular a amplitude total ( $A = 60 - 41 = 19$ )
- Determinar o número de classes  
 $k = \sqrt{n} = \sqrt{20} = 4,47$
- Calcular amplitude das classes  
 $c = \frac{A}{k-1} = \frac{19}{5-1} = 4,75 \cong 5$
- Calcular o limite inferior da primeira classe  
 $Li_1 = \text{menor valor} - c/2$
- Montar as classes e frequências

---

---

---

---

---

---

---

---

---

---



Tabela 3. Distribuição de frequências do tempo (seg) de início do atendimento em uma central de relacionamento com o cliente

Classes	Frequências absolutas
38,5  – 43,5	3
43,5  – 48,5	4
48,5  – 53,5	7
53,5  – 58,5	4
58,5  – 63,5	2
Total	20

---

---

---

---

---

---

---

---

---

---



- Classes Ex: 38,5 |– 43,5
- Limites de classes Ex: 38,5
- Número de classes Ex:  $k = 5$
- Amplitude ou intervalo de classe Ex: 5
- Amplitude total da amostra Ex: 19
- Ponto Médio da classe (H.T.B) Ex: 41
- Limite inferior da primeira classe Ex: 38,5

---

---

---

---

---

---

---

---

---

---



### Frequência acumulada

Tabela 4. Distribuição de frequências acumulada do tempo (min) de início de atendimento na central de relacionamento com o cliente.

Classes	Frequência acumulada para baixo
abaixo de 43,5	3
abaixo de 48,5	7
abaixo de 53,5	14
abaixo de 58,5	18
abaixo de 63,5	20

---

---

---

---

---

---

---

---

---

---



- A representação gráfica da variável contínua deve dar a idéia de continuidade
- Para frequências simples os tipos de gráficos mais utilizados são o histograma e o polígono de frequências
- Para as frequências acumuladas usa-se histogramas acumulados ou as ogivas (gráficos de frequências acumuladas).

---

---

---

---

---

---

---

---

---

---

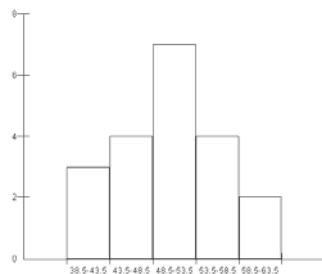


Figura 4. Histograma do tempo (seg) de início do atendimento dos clientes por uma central de relacionamento.

---

---

---

---

---

---

---

---

---

---

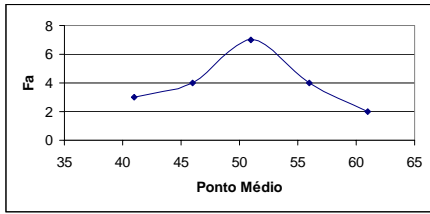


FIGURA 5. Polígono de frequências do tempo (seg) de início do atendimento dos clientes por uma central de relacionamento.

---

---

---

---

---

---

---

---

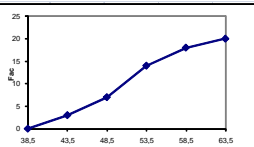


FIGURA 5 Ogiva "abaixo de" do tempo (seg) de início do atendimento dos clientes por uma central de relacionamento.

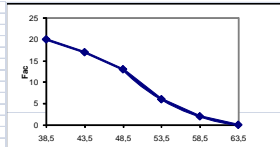


FIGURA 6 Ogiva "acima de" do tempo (seg) de início do atendimento dos clientes por uma central de relacionamento.

---

---

---

---

---

---

---

---



### Natureza da distribuição

- Simétrica** - 50% das observações estão acima da média e 50% estão abaixo
- assimétrica à direita** – forte concentração de dados à esquerda
- assimétrica à esquerda** – forte concentração de dados à direita
- multimodal** – vários picos de frequência

---

---

---

---

---

---

---

---



### Interpolação em distribuição de frequências

Quando os dados são apresentados apenas em uma distribuição de frequências e necessitamos estimar um valor diferente dos limites de classes usamos a interpolação linear simples.

*Exemplo: Utilizando a distribuição de frequências do gasto mensal (R\$), com produtos de beleza, de uma população de mulheres adultas, estimar: a) % de mulheres com gasto mínimo de 93%; b) % com gasto entre 15% e 30% do salário mínimo.*

---

---

---

---

---

---

---

---



Distribuição de frequências do gasto mensal (R\$), com produtos de beleza, por um grupo de mulheres

classes	frequência
30  --- 50	3
50  --- 70	8
70  --- 90	10
90  --- 110	18
110  --- 130	12
130  --- 150	7
150  --- 170	2
Total	60

---

---

---

---

---

---

---

---



## MEDIDAS DE POSIÇÃO

### Introdução

- **Objetivo** → Resumir o conjunto de dados em alguns valores que possam representar a variável
- **Principais medidas de posição** → Média aritmética, Mediana e a Moda

### Hipótese Tabular Básica (HTB)

Os dados contidos em uma determinada classe são representados pelo ponto médio da classe.

---

---

---

---

---

---

---

---



### Média aritmética

- A média aritmética é uma medida de posição que procura sintetizar um conjunto de observações.
- É muito utilizada: por apresentar:
  - apresenta facilidade de cálculo e de interpretação;
  - adapta-se bem à tratamentos algébricos
  - é um estimador que produz estimativas sem tendência, consistente, suficiente e geralmente, com boa precisão.
- Simbologia:  $\mu = \text{média populacional}$   
 $\bar{x} = \text{média amostral}$

---

---

---

---

---

---

---

---



Em Alguns casos ela não é considerada a melhor medida de posição para um conjunto de dados, por exemplo, em distribuições assimétricas como a distribuição de renda no Brasil a média aritmética não seria um valor que mostra a realidade da distribuição de renda no País.

A fórmula de cálculo da média com **dados não agrupados** é :

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

---

---

---

---

---

---

---

---



A fórmula de cálculo para **dados agrupados** (distribuição de freqüências) é:

$$\bar{x} = \frac{\sum X_i f_i}{\sum f_i} = \frac{\sum X_i f r_i}{\sum f r_i}$$

$X_i$  = ponto médio da classe  $i$ ;  $f_i$  a freqüência da classe  $i$

#### Exemplos de cálculo

1) A seguir temos tempos (min) que pessoas gastaram para realizar determinado teste.

30 50 80 30 100 80 70 110 70 60

Calcular o tempo médio.

---

---

---

---

---

---

---

---



2) Deseja saber o número médio de falhas por dia no equipamento X. A amostra foi composta de 5 observações.

2 1 1 3 0

3) Calcular o gasto médio com produtos de beleza (tabela do exemplo de interpolação)

**Propriedades da média aritmética**

- A soma algébrica dos desvios em relação à média é nula.

---

---

---

---

---

---

---

---

---

---



• A soma de quadrados dos desvios de um conjunto de dados, em relação a uma constante qualquer K, será mínima se e somente se  $K = \bar{X}$ .

• Somando-se (ou subtraindo-se) uma constante (c) a todos os valores de uma variável, a média do conjunto fica aumentada (ou diminuída) dessa constante.

• Multiplicando-se (ou dividindo-se) todos os valores de uma variável por uma constante (c), a média do conjunto fica multiplicada (ou dividida) por essa constante.

---

---

---

---

---

---

---

---

---

---



**Outros tipos de médias**

Ponderada => 
$$\bar{x} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

Geométrica => 
$$Mg = \sqrt[n]{x_1^{F_1} \times x_2^{F_2} \times \dots \times x_n^{F_n}}$$

Harmônica => 
$$Mh = \frac{n}{\frac{F_1}{x_1} + \frac{F_2}{x_2} + \dots + \frac{F_n}{x_n}} = \frac{n}{\sum_{i=1}^n \frac{F_i}{x_i}}$$

---

---

---

---

---

---

---

---

---

---



**Exemplos:**

- 1) As notas obtidas em um concurso nas disciplinas A, B, C, D, foram, respectivamente, 80; 85; 90 e 70. Se os pesos destas disciplinas são: 3, 5, 2, 1, respectivamente, determine a média ponderada.
- 2) Suponha que o salário de 15 funcionários de uma empresa seja: 1 1 1 1 1 1 1 2 2 2 2 3 4 10 20. Determine a média aritmética e a média geométrica, destes dados.
- 3) Suponha que em um experimento em que o pesquisador avaliou o preço mensal de um determinado produto ele tenha encontrado: 5 8 10 6 8 11 5 7 9 4 7 10. Calcular a média aritmética e a média harmônica.

---

---

---

---

---

---

---

---

---

---

---

---



**Mediana (Md)**

- Realização (valor) que ocupa a posição central de um conjunto de dados ordenados.
- Se o número de observações for ímpar a mediana será o valor central.
- Se o número de dados for par a mediana será igual a média aritmética entre os dois valores centrais do conjunto de dados.

---

---

---

---

---

---

---

---

---

---

---

---



**Ex. 1.** Juros cobrados em certa operação financeira por diferentes instituições. Determinar Md.

1,8	1,9	1,7	1,8	2,1	2,0	2,4	2,2
1,8	2,0	1,9					

**Ex. 2.** Número de vendas do produto X ao longo do ano por um revendedor. Determinar a mediana.

28	27	30	25	22	15	14	21	25
26	29	29						

---

---

---

---

---

---

---

---

---

---

---

---



- Para os **dados agrupados**:

$$Md = Li + \left( \frac{(n/2) - F_{antac}}{f_{med}} \right) \times c$$

Li = limite inferior da classe mediana

F<sub>antac</sub> = freq. Acumulada anterior a Md

f<sub>med</sub> = freq. da classe mediana

**Opção alternativa** → usar o ponto médio da classe mediana (*Hipótese Tabular Básica*)

---

---

---

---

---

---

---

---



**Ex.:** Determinar a mediana para a taxa de peças a serem remanufaturadas.

Tabela 5. Taxa de peças a serem remanufaturadas em uma linha de produção

Classes	Freq.	freq. Acum.
9,9  – 18,62	10	10
18,62  – 27,34	13	23
27,34  – 36,06	6	29
36,06  – 44,78	4	33
44,78  – 53,5	0	33
53,5  – 62,2	1	34
<b>Total</b>	<b>34</b>	<b>-</b>

---

---

---

---

---

---

---

---



### Moda

- Valor que ocorre com maior freqüência entre os valores observados
- Um conjunto de observações pode ter uma única moda, mais de uma moda ou não ter nenhum valor modal.

**Exemplo:** Para os conjuntos de dados abaixo, encontre o valor modal.

1) Produção de leite (kg/dia) das vacas de pequeno produtor rural.

8 8 9 10 11 11 11 12 12 12 12 12 13  
13 13 13 13 14 15 15

---

---

---

---

---

---

---

---



2) Número de vezes que clientes usaram o cartão de crédito no último mês.

7 8 5 11 13 9 6 12 10 15

3) notas obtidas por 15 alunos em determinada disciplina.

6 4 7 7 8 8 6 7 9 7 8 7 7 5 7

---

---

---

---

---

---

---

---



• Para os **dados agrupados**:

$$Mo = L_i + \frac{d_1}{d_1 + d_2} \times c$$

Li = limite inferior da classe modal  
d<sub>1</sub> = dif entre freq das classes modal e anterior  
d<sub>2</sub> = dif entre freq das classes modal e posterior

*Opção alternativa → usar o ponto médio da classe mediana (Hipótese Tabular Básica)*

**Exemplo:** Determinar a moda para o exemplo das taxas de peças remanufaturadas

---

---

---

---

---

---

---

---



**RELAÇÃO ENTRE MÉDIA, MEDIANA E MODA E O TIPO DE DISTRIBUIÇÃO**

- a) Se  $\bar{X} = Md = Mo \rightarrow$  distribuição simétrica
- b) Se  $Mo < Md < \bar{X} \rightarrow$  distribuição assimétrica a direita
- c) Se  $\bar{X} < Md < Mo \rightarrow$  distribuição assimétrica a esquerda

---

---

---

---

---

---

---

---



Qual medida posição devo utilizar para expressar uma determinada variável?

**Ferramentas diferentes para situações diferentes**




---

---

---

---

---

---

---

---



**Separatrizes (quartil, decil, percentil)**

- **Quartil** → separação dos dados em 4 partes iguais
  - **Decil** → separação dos dados em 10 partes iguais
  - **Percentil** → separação dos dados em 100 partes iguais
- Quartis:**  $Q_1$  → primeiro quartil;  $Q_2$  → segundo quartil e  $Q_3$  → terceiro quartil
- Decis:**  $D_1$  → primeiro decil; ...;  $D_9$  → nono decil
- Percentis:**  $P_1$  → primeiro percentil; ...;  $P_{99}$  → percentil 99

---

---

---

---

---

---

---

---



- Note que, abaixo de  $Q_1$  temos 25% dos dados, entre  $Q_1$  e  $Q_2$  temos 25%, entre  $Q_2$  e  $Q_3$  temos 25% e acima de  $Q_3$  também 25%. De maneira análoga pode-se fazer com decil ( 10% para cada decil) e percentil (1% para cada percentil).
- Note também que:  $Q_1 = P_{25}$ ;  $Q_2 = P_{50} = Md = D_5$ ;  $Q_3 = P_{75}$
- Para o caso de distribuição de frequências acumuladas pode-se adotar o critério da definição das separatrizes assumindo o ponto médio das respectivas classes.

---

---

---

---

---

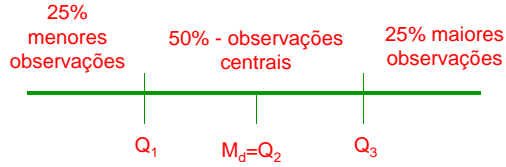
---

---

---

## Quartis

### Amostra ordenada



$Q_1$ : primeiro quartil

$Q_2$ : segundo quartil (mediana)

$Q_3$ : terceiro quartil

---

---

---

---

---

---

---

---

**Exemplo:** Determinar os quartis para:

- Idade em anos de pessoas aprovadas no vestibular  
18 19 17 18 21 20 24 22  
18 20 19
- Dados de temperatura mínima mensal (°C) em uma certa região.  
28 27 30 25 22 15 14 21 25  
26 29 29
- Número de gols por partida em um certo campeonato de futebol

Distribuição de freqüências do número de gols por partida

Número de trincas	fa
0	125
1	45
2	15
3	10
4	4
5	1
<b>Total</b>	<b>200</b>

---

---

---

---

---

---

---

---

**Calcular a média, a mediana e a moda de cada uma das amostras abaixo (Número de Ações Negociadas)**

A	B
56	33
56	42
57	48
58	52
61	57
63	67
63	67
67	77
67	82
67	90

---

---

---

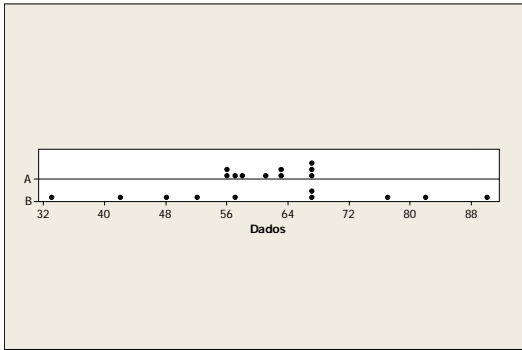
---

---

---

---

---




---

---

---

---

---

---

---

---

### MEDIDAS DE DISPERSÃO

- As medidas de posição → não apresentam informações sobre o comportamento da variável.
- Necessário associar à medida de posição, uma medida de dispersão (variabilidade).

---

---

---

---

---

---

---

---

• *Exemplo* → média é medida de centro da distribuição, porém, nada informa com relação a dispersão dos valores. Quanto maior a variação dos dados menor a representatividade da média. Assim, dizemos que as medidas de dispersão servem para qualificar a média. Quanto menor a dispersão, mais confiável é a média.

- As principais medidas de dispersão são: **Amplitude total (A)**; **variância ( $s^2$ )** e **desvio padrão (s)**; **coeficiente de variação (CV)**; **erro padrão da média (S)**.

---

---

---

---

---

---

---

---





· alternativa - módulo

$$|d| = \sum_{i=1}^n (x_i - \bar{x})$$

Daí vem a definição de **Desvio médio**

- evitar que amostras de tamanhos diferentes, que possuíssem mesma dispersão, apresentassem desvios diferentes, ponderou-se o modulo dos desvios por n

$$|DM| = \frac{\sum_{i=1}^n (x_i - \bar{x})}{n}$$

---

---

---

---

---

---

---

---



*Problemas com essa medida:*

- dificuldade de cálculo e de interpretação em algumas situações
- Não é uma medida mínima de erro

*Outra proposta - trabalhar com quadrados*

**Desvio quadrático médio**

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

---

---

---

---

---

---

---

---



*Vantagens da medida*

- É uma média mínima de erro
- Apresenta facilidade de cálculo e de interpretação
- Apresenta boas propriedades algébricas

*Desvantagem*

- A medida de dispersão definida como acima apresenta o inconveniente de ser tendenciosa, ou seja, de afastar do valor real da população.

---

---

---

---

---

---

---

---



### Como contornar ?

- Retirar tendência → ver propriedades dos estimadores (inferência estatística)
- Surge a definição de **variância** e **desvio padrão** → medidas de dispersão amplamente utilizadas na estatística
- Grande importância na caracterização de variáveis e nas inferências paramétricas.

---

---

---

---

---

---

---

---



### Variância e desvio padrão

#### Variância

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

**Desvio padrão** => raiz quadrada positiva da **variância**

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

---

---

---

---

---

---

---

---



**Exemplo 1:** Os dados a seguir representam o número de atendimentos de emergência em uma unidade de saúde em uma semana.

22 15 15 18 17 26 25

Determinar a média e o desvio padrão e interpretar.

#### Dados agrupados em classes

$$s^2 = \frac{SQD}{(\sum f_i) - 1} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2 f_i}{(\sum f_i) - 1} = \frac{\sum X_i^2 f_i - \frac{(\sum X_i f_i)^2}{\sum f_i}}{(\sum f_i) - 1}$$

$$s = +\sqrt{s^2}$$

---

---

---

---

---

---

---

---



*Exemplo:* Determinar a média, a variância e o desvio padrão dos gastos com produtos de beleza

Distribuição de frequências do gasto mensal (R\$), com produtos de beleza, por um grupo de mulheres

classes	frequência
30  --- 50	3
50  --- 70	8
70  --- 90	10
90  --- 110	18
110  --- 130	12
130  --- 150	7
150  --- 170	2
Total	60

---

---

---

---

---

---

---

---

---

---



### Propriedades e características

- O desvio padrão é um valor mínimo de erro, pois os desvios são calculados em relação à média aritmética
- São estatísticas que utilizam todas as observações no cálculo e também sofrem pouca influência de mudanças amostrais
- Valores extremos exercem maior influência que os valores centrais (próximos à média)

---

---

---

---

---

---

---

---

---

---



- Variância e desvio padrão são sempre positivos
- São estatísticas de grande importância na descrição das populações e nas inferências
- A variância de uma constante é sempre igual a zero
- A variância de uma soma de variáveis será igual a soma das variâncias de cada variável, se e somente se as variáveis forem independentes

---

---

---

---

---

---

---

---

---

---



- Se somarmos ou subtraírmos uma constante a cada valor observado e calcularmos a nova variância e o novo desvio padrão, verificaremos que estes não se alteram em relação aos dados iniciais.
- A multiplicação ou divisão de cada observação por uma constante, faz com que a nova variância seja multiplicada ou dividida pela constante ao quadrado e novo desvio padrão fique multiplicado ou dividido pela constante.

---

---

---

---

---

---

---

---



*Exemplo:* Suponha que o salário médio de uma empresa X é de R\$ 1500,00 com desvio padrão de R\$ 250,00. Será dado um aumento de 10% e um abono de R\$ 50,00 para cada funcionário. Qual o novo salário médio e o novo desvio padrão?

*Obs:* O desvio padrão é uma medida absoluta de variabilidade de um atributo. Essa medida não deve ser utilizada na comparação de variabilidade entre amostras ou entre variáveis.

*Como comparar variabilidade??*

---

---

---

---

---

---

---

---



*Exemplo:* Qual dos atributos apresentados abaixo tem maior variabilidade?

	Temp (°C)	Prec (mm)
Média	22	800
s	5	100

### COEFICIENTE DE VARIAÇÃO

Medida de dispersão usada para comparação de variabilidade

$$CV = 100 \cdot \frac{s}{\bar{X}}$$

---

---

---

---

---

---

---

---



### ERRO PADRÃO DA MÉDIA

Expressa a precisão com que a média foi estimada

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

**Exemplo:** Determinar o erro padrão para o tempo de espera na fila do banco, usando uma amostra de tamanho 7 e uma amostra de tamanho 15

(A) 20 15 15 18 17 16 15

(B) 15 18 15 18 17 20 22 14 16 15 15 17 19 16 15

---

---

---

---

---

---

---

---

---

---



### CARACTERIZAÇÃO DE OUTLIERS E VALORES EXTREMOS

- Outliers e valores extremos são aqueles que estão “muito” afastados do centro da distribuição.
- Uma forma de caracterizá-los é através do desenho esquemático (Box-Plot)
- BOX – PLOT → Caixa que contém 50% da distribuição. Parte superior é  $Q_3$  e o inferior é  $Q_1$ . Mediana  $Q_2$  está dentro da caixa.
- Desvio interquartilício (H):

$$H = Q_3 - Q_1$$

---

---

---

---

---

---

---

---

---

---



Valores não discrepantes

$$Q_1 - 1,5H \leq X_i \leq Q_3 + 1,5H$$

Para a análise de outliers, geralmente, considera-se:

$$Q_3 + 1,5H \leq X_i \leq Q_3 + 3H$$

$$Q_1 - 3H \leq X_i \leq Q_1 - 1,5H$$

Para a análise de valores extremos, geralmente, considera-se:

$$X_i > Q_3 + 3H$$

$$X_i < Q_1 - 3H$$

---

---

---

---

---

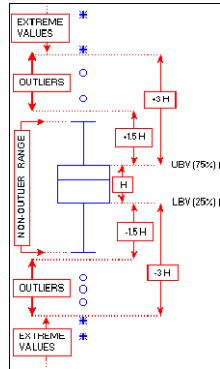
---

---

---

---

---




---

---

---

---

---

---

---

---



### Um exemplo

Uma empresa suspeita que os fornecedores de um certo componente, com fábrica num determinado município (A), estejam fazendo uma política combinada de preços (cartel). Para verificar essa acusação, foram tomados os preços praticados por uma amostra de 20 fábricas desse município e de 25 fábricas de municípios vizinhos (Controle).

---

---

---

---

---

---

---

---



### Dados

Município			
Controle		A	
14,80	12,90	21,30	20,30
18,20	20,90	20,70	19,60
13,60	19,30	20,70	19,20
15,50	14,40	19,90	18,50
12,00	15,10	20,30	18,60
13,70	13,10	21,10	20,30
16,00	15,50	19,60	20,10
17,30	14,30	19,30	19,90
14,40	15,10	20,80	21,00
16,10	15,80	19,70	18,90
26,80	13,00		
12,10	14,90		
	17,00		

---

---

---

---

---

---

---

---



### ASSIMETRIA E CURTOSE

O coeficiente de assimetria mostra o afastamento da variável em relação a um valor central, ou seja, na distribuição simétrica tem-se 50% dos valores observados acima da observação central e 50% abaixo. Se a distribuição é assimétrica, esta relação não é observada.

O coeficiente de curtose mostra a dispersão (achatamento) da distribuição em relação a um padrão, geralmente a curva normal.

Estes dois coeficientes são utilizados para inferências sobre a normalidade da variável em estudo.

---

---

---

---

---

---

---

---



### ASSIMETRIA

**Assimetria:** significa desvio ou afastamento da simetria, (grau de deformação de uma curva).

# **Simétrica**, se a média e a moda coincidem.

# **Assimétrica à esquerda ou negativa**, se a média é menor que a moda.

# **Assimétrica à direita ou positiva**, se a média é maior que a moda.




---

---

---

---

---

---

---

---



### Como avaliar a assimetria de uma variável

• Uma primeira idéia pode ser pela comparação da média, da mediana e da moda

$$\bar{X} = Md = Mo$$

• Podemos também comparar os quartis: **Q3-Md** e **Md-Q1**

Note que estas medidas de assimetria apresentam valores absolutos, elas não permitem a comparação entre as medidas de duas distribuições, da mesma forma que ocorreu com o desvio padrão.

---

---

---

---

---

---

---

---



Exemplo:

Distribuição A

Número de ligações perdidas	$f_i$
2   6	6
6   10	12
10   14	24
14   18	12
18   22	6
	60

$\bar{x} = 12$   
Md = 12  
Mo = 12  
s = 4,42

Distribuição B

Número de ligações perdidas	$f_i$
2   6	6
6   10	12
10   14	24
14   18	30
18   22	6
	78

$\bar{x} = 12,9$   
Md = 13,5  
Mo = 16  
s = 4,2

Distribuição C

Número de ligações perdidas	$f_i$
2   6	6
6   10	30
10   14	24
14   18	12
18   22	6
	78

$\bar{x} = 11,1$   
Md = 10,5  
Mo = 8  
s = 4,2

---

---

---

---

---

---

---

---

---

---

---

---



Recomenda-se trabalhar com medidas adimensionais com o objetivo de facilitar as comparações.

**Primeiro coeficiente de assimetria de Pearson**

$$As_1 = \frac{\bar{X} - Mo}{s}$$

Não é muito utilizado, pois o conjunto de dados pode não apresentar valor modal

**Segundo coeficiente de assimetria de Pearson**

$$As = \frac{3(\bar{x} - Md)}{s}$$

Se  $0,15 < |As| < 1$ , a assimetria é moderada.  
Se  $|As| > 1$ , a assimetria é forte.

---

---

---

---

---

---

---

---

---

---

---

---



Para o exemplo do número de ligações perdidas temos:

$$As_A = \frac{3(12 - 12)}{4,42} = 0 \Rightarrow \text{simetria}$$

$$As_B = \frac{3(12,9 - 13,5)}{4,2} = -0,429 \Rightarrow \text{assimetria negativa}$$

$$As_C = \frac{3(11,1 - 10,5)}{4,2} = +0,429 \Rightarrow \text{assimetria positiva}$$

Estas são algumas formas de se avaliar a assimetria de um conjunto de dados, entretanto o coeficiente mais utilizado se baseia no terceiro **momento centrado na média**

---

---

---

---

---

---

---

---

---

---

---

---



### Momento Estatístico

Se  $x_1, x_2, \dots, x_n$  são os  $n$  valores assumidos pela variável  $X$ , definimos o momento de ordem  $t$  dessa variável como:

$$M_t = \frac{\sum_{i=1}^n x_i^t}{n}$$

Note que se  $t=1$  temos a média aritmética, ou seja, a média aritmética é igual ao primeiro momento em relação à origem.

O momento de ordem  $t$  centrado em uma constante  $K$ , com  $K \neq 0$  é definido como:

$$M_t^K = \frac{\sum_{i=1}^n (x_i - K)^t}{n}$$

---

---

---

---

---

---

---

---

---

---

---

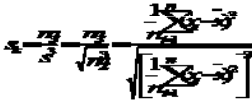
---



### Observações:

- se  $t = 1$  e  $K = \bar{X}$ , temos  $M_1^{\bar{X}} = m_1 = 0 \rightarrow$  (propriedade da média aritmética)
- se  $t=2$  e  $K = \bar{X}$ , temos  $M_2^{\bar{X}} = m_2 = \sigma^2$ .

### Coefficiente do momento de assimetria



- $s_k = 0$ , se houver simetria
- $s_k > 0$ , se houver assimetria à direita
- $s_k < 0$ , se houver assimetria à esquerda

---

---

---

---

---

---

---

---

---

---

---

---



### CURTÓSE

Mede o grau de achatamento ou afilamento de uma distribuição em relação a uma distribuição padrão, denominada **curva normal**.

#### Classificação da distribuição de acordo com a curtose

- **mesocúrtica**, se a curva de freqüências apresentar um grau de achatamento equivalente ao da curva normal.
- **leptocúrtica**, se a curva de freqüências apresenta-se mais fechada (ou mais afilada na parte superior) que a curva normal.
- **platicúrtica**, se a curva de freqüências apresenta-se mais aberta (ou mais achatada na parte superior) que a curva normal.

---

---

---

---

---

---

---

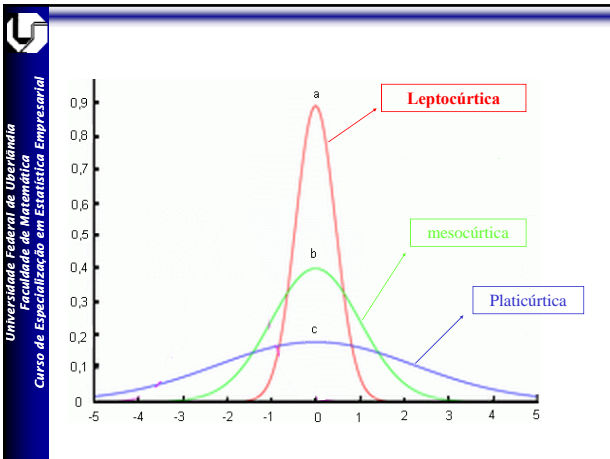
---

---

---

---

---




---

---

---

---

---

---

---

---

Universidade Federal de Uberlândia  
Faculdade de Matemática  
Curso de Especialização em Estatística Empresarial

**Como avaliar a curtose de uma variável**

**Coefficiente percentílico de curtose:**

$$C = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$$

Curva normal → C = 0,263

C = 0,263 ⇒ curva mesocúrtica  
C < 0,263 ⇒ curva leptocúrtica  
C > 0,263 ⇒ curva platicúrtica

---

---

---

---

---

---

---

---

Universidade Federal de Uberlândia  
Faculdade de Matemática  
Curso de Especialização em Estatística Empresarial

**Coefficiente do momento de curtose**

$$k = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^4}{\left[ \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right]^2}$$

$$\begin{cases} k = 3, & \text{meso} \\ k > 3, & \text{lepto} \\ k < 3, & \text{plati} \end{cases}$$

Mais utilizado na maioria dos softwares estatísticos

**Observação:**

Em alguns programas computacionais como o Excel, Statistica e GS+ existe uma padronização do valor de k e o valor de comparação é o zero, portanto, se k = 0 → mesocúrtica, se k < 0 → platicúrtica e se k > 0 → leptocúrtica.

---

---

---

---

---

---

---

---



### Erro Padrão dos Coeficientes de Assimetria e de Curtose

Para uma melhor interpretação do coeficiente de assimetria e do coeficiente de curtose, alguns programas, fornecem também o erro padrão desses coeficientes e a partir dos valores dos coeficientes associados com seus respectivos erros padrão, pode-se concluir se os dados tem distribuição normal ou não

Estes erros padrão dependem da distribuição, mas podem ser aproximados por:

$$s_{s_k} = \sqrt{\frac{6}{n}} \text{ erro padrão de } S_k$$

$$s_k = \sqrt{\frac{24}{n}} \text{ erro padrão de } k$$

---

---

---

---

---

---

---

---

---

---

---

---



Por exemplo: Se o valor obtido na amostra para  $S_k = 0,30$  com erro padrão de  $0,65$  e se o valor de  $k = 2,5$  com erro padrão de  $0,80$ , podemos dizer que a distribuição tende a normal (simétrica e mesocúrtica), pois  $0,3 \pm 0,65$  e  $2,5 \pm 0,80$ , incluem os valores zero e três, respectivamente.

#### Exemplo de aplicação

Determinar os coeficientes de assimetria e de curtose para:

1) Número de vendas por dia do produto X.

10	15	12	18	22	15	15	13
11	15	15	12	17	16	12	15
13	13	20	19	18			

---

---

---

---

---

---

---

---

---

---

---

---



#### Solução:

X	(x-media)	(x-media) <sup>2</sup>	(x-media) <sup>3</sup>	(x-media) <sup>4</sup>		
10.0000	-5.0476	25.4785	-128.6056	649.1518		
15.0000	-0.0476	0.0023	-0.0001	0.0000		
12.0000	-3.0476	9.2880	-28.3062	86.2666		
18.0000	2.9524	8.7166	25.7346	75.9783		
22.0000	6.9524	48.3356	336.0475	2336.3303		
15.0000	-0.0476	0.0023	-0.0001	0.0000	Cs=	0.4675
15.0000	-0.0476	0.0023	-0.0001	0.0000	Ck=	2.5824
13.0000	-2.0476	4.1927	-8.5851	17.5791		
11.0000	-4.0476	16.3832	-66.3130	268.4099		
15.0000	-0.0476	0.0023	-0.0001	0.0000		
15.0000	-0.0476	0.0023	-0.0001	0.0000		
12.0000	-3.0476	9.2880	-28.3062	86.2666		
17.0000	1.9524	3.8118	7.4421	14.5298		
16.0000	0.9524	0.9070	0.8638	0.8227		
12.0000	-3.0476	9.2880	-28.3062	86.2666		
15.0000	-0.0476	0.0023	-0.0001	0.0000		
13.0000	-2.0476	4.1927	-8.5851	17.5791		
13.0000	-2.0476	4.1927	-8.5851	17.5791		
20.0000	4.9524	24.5261	121.4626	601.5285		
19.0000	3.9524	15.6213	61.7414	244.0255		
18.0000	2.9524	8.7166	25.7346	75.9783		
media	15.0476					
soma	0.0000	192.9524	273.4331	4578.2922		
momento	0.0000	9.1882	13.0206	218.0139		

---

---

---

---

---

---

---

---

---

---

---

---



2) tempo gasto na tarefa

Distribuição de freqüências tempo (min) gasto na tarefa X.

Tempo (min)	f
20  ----25	8
25  ----30	12
30  ----35	20
35  ----40	8
40  ----45	2

---

---

---

---

---

---

---

---

---

---



Tempo (min)	f	xi	(x-media).f	(x-media) <sup>2</sup> .f	(x-media) <sup>3</sup> .f	(x-media) <sup>4</sup> .f
20  ----25	8	22.5	-67.2	564.48	-4741.632	33829.7088
25  ----30	12	27.5	-40.8	138.72	-471.648	1603.6032
30  ----35	20	32.5	32	51.2	61.92	131.072
35  ----40	8	37.5	52.8	348.48	2299.968	15179.7888
40  ----45	2	42.5	23.2	269.12	3121.792	36212.7872
		50				
media =	30.9					
soma			7.10543E-14	1372	290.4	92956.96
momento			1.42109E-15	27.44	5.808	1889.1392
Cs	0.040406427					
Ck	2.469128509					

---

---

---

---

---

---

---

---

---

---