



Centro Federal de Educação Tecnológica Celso Suckow da Fonseca
Campus Maracanã
Departamento de Informática

Estatística e Probabilidade

Prof: Pedro Henrique González

1

Estatística – Introdução

A Estatística subdivide-se em três áreas: estatística descritiva, probabilidade e inferência estatísticas.

Estatística Descritiva:

- A estatística descritiva, como o próprio nome já diz, se preocupa em descrever os dados.

Probabilidade:

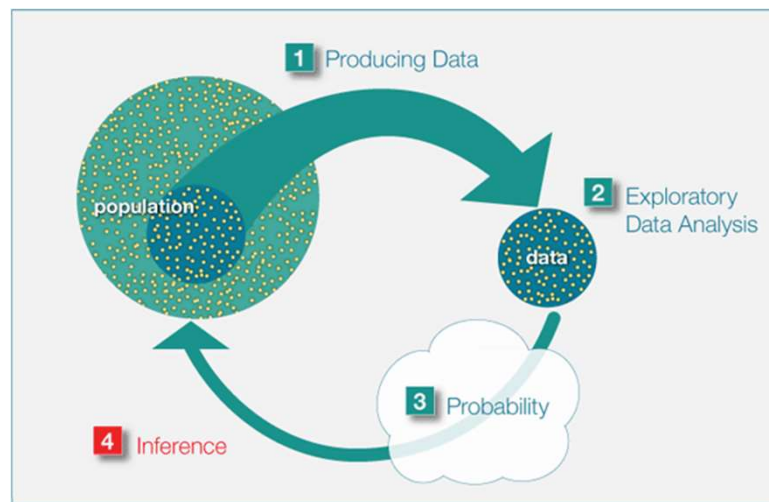
- A probabilidade nos permite descrever os fenômenos aleatórios, ou seja, aqueles em que esta presente a incerteza.

Inferência Estatística :

- A inferência estatística, fundamentada na teoria das probabilidades, se preocupa com a análise destes dados e sua interpretação.

2

E como se relacionam?



<https://jailsonr.medium.com/o-problema-da-infer%C3%Aancia-estat%C3%ADstica-661bff138130>

3

Organização da Disciplina

- Estatística Descritiva
- Introdução a Probabilidade
- Modelos Discretos Univariados
- Modelos Contínuos Univariados
- Valor Esperado e Desvio Padrão
- Dependência entre Variáveis Aleatórias
- Distribuições Conjuntas
- Distribuição Marginal
- Probabilidade Condicional

4

Agenda

- Estatística Descritiva
- Introdução a Probabilidade
- Modelos Discretos Univariados
- Modelos Contínuos Univariados
- Valor Esperado e Desvio Padrão
- Dependência entre Variáveis Aleatórias
- Distribuições Conjuntas
- Distribuição Marginal
- Probabilidade Condicional

5

Estatística Descritiva

- Por que é importante?

Pois nos fornece as ferramentas para descrever dados!

- Mas quais ferramentas?

1. Tabelas
2. Gráficos
3. Medidas Descritivas

6

Estatística Descritiva: Variável

Antes de continuarmos...

Variável: é a característica que vai ser observada, medida ou contada nos elementos da população ou da amostra e que pode variar, ou seja, assumir um valor diferente de elemento para elemento.

As variáveis podem ser divididas em: **Qualitativa e Quantitativa**

Variável Qualitativa: é uma variável que assume como possíveis valores atributos ou qualidades. Também são denominadas variáveis categóricas.

Variável Quantitativa: é uma variável que assume como possíveis valores, números.

7

Estatística Descritiva: Variável Qualitativa

Antes de entrarmos de continuarmos...

As variáveis podem ser divididas em: **Nominal e Ordinal**

Variável Qualitativa Nominal: é uma variável que assume como possíveis valores atributos ou qualidades e estes não apresentam uma ordem natural de ocorrência.

Variável Qualitativa Ordinal: é uma variável que assume como possíveis valores atributos ou qualidades e estes apresentam uma ordem natural de ocorrência.

8

Estatística Descritiva: Variável Quantitativa

Antes de entrarmos de continuarmos...

As variáveis podem ser divididas em: **Discreta e Contínua**

Variável Quantitativa Discreta: é e uma variável que assume como possíveis valores números, em geral inteiros, formando um conjunto finito ou enumerável.

Variável Quantitativa Contínua : é uma variável que assume como possíveis valores números, em intervalos da reta real e, em geral, resultantes de mensurações.

9

Estatística Descritiva: Exemplos de Variáveis

Classifique os exemplos a seguir:

1. Numero de reprovações, por disciplina, dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: 0; 1; 2;
2. Estado civil dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: solteiro, casado, separado.

10

Estatística Descritiva: Exemplos de Variáveis

Classifique os exemplos a seguir:

1. Numero de reprovações, por disciplina, dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: 0; 1; 2;

R: Variável quantitativa discreta

2. Estado civil dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: solteiro, casado, separado.

11

Estatística Descritiva: Exemplos de Variáveis

Classifique os exemplos a seguir:

1. Numero de reprovações, por disciplina, dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: 0; 1; 2;

R: Variável quantitativa discreta

2. Estado civil dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: solteiro, casado, separado.

R: Variável qualitativa ordinal

12

Estatística Descritiva: Exemplos de Variáveis

Classifique os exemplos a seguir:

3. Peso (quilogramas) dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: 58, 59, 63,...
4. Meios de informação utilizados pelos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: televisão, revista, internet, jornal.

13

Estatística Descritiva: Exemplos de Variáveis

Classifique os exemplos a seguir:

3. Peso (quilogramas) dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: 58, 59, 63,...
- R:** Variável quantitativa contínua
4. Meios de informação utilizados pelos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: televisão, revista, internet, jornal.

14

Estatística Descritiva: Exemplos de Variáveis

Classifique os exemplos a seguir:

3. Peso (quilogramas) dos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: 58, 59, 63,...

R: Variável quantitativa contínua

4. Meios de informação utilizados pelos alunos da disciplina Estatística do curso de Ciência da Computação do Cefet/RJ: televisão, revista, internet, jornal.

R: Variável qualitativa nominal

15

Estatística Descritiva: Tabelas

Quais tipos de tabela estudaremos?

Todas as variáveis podem ser resumidas através de uma tabela, mas a construção é diferenciada dependendo do tipo de variável.

Denominamos **tabela simples** a tabela que resume os dados de uma única variável qualitativa e **distribuição de frequências** ao resumo de uma única variável quantitativa.

16

Estatística Descritiva: Tabelas

Tabelas: Simples

Uma tabela simples contém as diferentes categorias observadas de uma **variável qualitativa** e suas respectivas contagens, denominadas **frequências absolutas**.

Classificação de uma Tabela Simples:

- **Temporal:** quando as observações são feitas levando-se em consideração o tempo;
- **Geográfica:** quando os dados referem-se ao local de ocorrência;
- **Específica (ou categórica):** quando tempo e local são fixos; e
- **Comparativa:** quando a tabela resume informações de duas ou mais variáveis. A tabela comparativa é também denominada tabela cruzada ou de dupla ou mais entradas.

17

Estatística Descritiva: Tabelas

Tabelas: Simples - Exemplo

Exemplo de uma tabela histórica

Tabela 02 – Número de alunos matriculados na disciplina Probabilidade I do curso de Estatística da Universidade Estadual de Maringá.

Ano	Nº de Alunos
2000	40
2001	59
2002	63
2003	69
2004	71

Fonte: DES/UEM.

Nota: Os números de 2003 e 2004 correspondem a duas turmas.

Exemplo de uma tabela geográfica, específica e comparativa

Tabela 03 – Município de procedência dos alunos da disciplina Inferência Estatística do curso de Estatística da Universidade Estadual de Maringá, 21/03/2005.

Município de Procedência	Nº de Alunos
Maringá	12
Outro no Paraná	7
Fora do Paraná	3
Total	22

Fonte: DES/UEM.

18

Estatística Descritiva: Tabelas

Tabelas: Frequência Relativa e/ou Relativa Percentual

É comum e útil na interpretação de tabelas a inclusão de uma coluna contendo as frequências relativas e/ou relativas em percentual.

A **frequência relativa** é obtida dividindo-se a frequência absoluta de cada categoria da variável pelo número total de observações (número de elementos da amostra ou da população).

Multiplicando este resultado por 100, obtemos a **frequência relativa em percentual**.

19

Estatística Descritiva: Tabelas

Tabelas: Frequência Relativa e/ou Relativa Percentual - Exemplo

Tabela 04 – Município de procedência dos alunos da disciplina Inferência Estatística do curso de Estatística da Universidade Estadual de Maringá, 21/03/2005.

Município de Procedência	Nº de Alunos	Percentual
Maringá	12	55
Outro no Paraná	7	32
Fora do Paraná	3	13
Total	22	100

Fonte: DES/UEM.

20

Estatística Descritiva: Tabelas

Tabelas: Frequência Acumulada

A frequência absoluta acumulada, denotada por F_{ai} , é obtida somando-se a frequência absoluta do valor considerado as frequências absolutas anteriores a este mesmo valor.

A frequência relativa acumulada, denotada por $f_{ai}\%$, é definida como:

$$f_{ai} \% = \frac{F_{ai}}{n} \times 100$$

21

Estatística Descritiva: Tabelas

Tabelas: Distribuição de Frequências Completa

Uma tabela contendo todas estas frequências é dita uma distribuição de frequências completa. Desta forma, a Tabela 8 pode ser apresentada como:

Tabela 08 – Número de irmãos dos alunos da disciplina Inferência Estatística do curso de Estatística da Universidade Estadual de Maringá, 21/03/2005.

Número de irmãos (x_i)	F_i	$f_i\%$	F_{ai}	$f_{ai}\%$
0	1	4,55	1	4,55
1	6	27,26	7	31,81
2	12	54,55	19	86,36
3	2	9,09	21	95,45
6	1	4,55	22	100,00
Total	22	100,00		

Fonte: DES/UEM.

22

Estatística Descritiva: Tabelas

Tabelas: Distribuição de Frequência em Classes com Perda de Informação

A **distribuição de frequências em classes** é apropriada para apresentar dados quantitativos contínuos ou discretos com um numero elevado de possíveis valores.

É necessário dividir os dados em intervalos ou faixas de valores, que são denominadas **classes**.

Uma **classe** é uma linha da distribuição de frequências. O menor valor da classe é denominado **limite inferior** (l_i) e o maior valor da classe é denominado **limite superior** (L_i).

23

Estatística Descritiva: Tabelas

Tabelas: Distribuição de Frequência em Classes com Perda de Informação Intervalos

O intervalo (ou classe) pode ser representado das seguintes maneiras:

1. l_i |----- L_i , em que o limite inferior da classe é incluído na contagem da frequência absoluta, mas o superior não;
2. l_i -----| L_i , em que o limite superior da classe é incluídos na contagem, mas o inferior não;
3. l_i |-----| L_i , em que tanto o limite inferior quanto o superior são incluídos na contagem;
4. l_i ----- L_i , em que os limites não fazem parte da contagem.

Podemos escolher qualquer uma destas opções. O importante é tornar claro no texto ou na tabela qual está sendo usada.

24

Estatística Descritiva: Tabelas

Tabelas: Distribuição de Frequência em Classes com Perda de Informação Número de Classes

Geralmente, os seguintes critérios são utilizados para a determinação do **número de classes**, denotado por k :

- Raiz quadrada: $k = \sqrt{n}$,
- log (Sturges): $k = 1 + 3,3 \log(n)$,
- ln (Milone): $k = -1 + 2 \ln(n)$,
- $k = 1 + 10^d AT$,

em que n é o número de elementos da amostra, AT é a **amplitude total** dos dados e d é o número de decimais de seus elementos.

Devemos lembrar que, sendo k o número de classes, o resultado obtido por cada um dos critérios deve ser o número inteiro mais próximo ao obtido.

25

Estatística Descritiva: Tabelas

Tabelas: Distribuição de Frequência em Classes com Perda de Informação Amplitude de Cada Classes

Determinado o número de classes da distribuição de frequências, o próximo passo é determinar a amplitude de cada classe, h , definida por:

$$h = \frac{AT}{k}.$$

Portanto, todas as classes terão a mesma amplitude, o que permitirá a construção de gráficos e o cálculo de medidas descritivas.

26

Estatística Descritiva: Tabelas

Tabelas: Distribuição de Frequência em Classes com Perda de Informação Pontos Médios

No caso de uma distribuição de frequência contínua, ou em classes, uma outra coluna pode ser acrescentada à tabela. É a coluna dos **pontos médios**, denotada por x_i e definida como a média dos limites da classe:

$$x_i = \frac{l_i + L_i}{2}; \quad i = 1 \dots k.$$

Estes valores são utilizados na construção de gráficos e na obtenção de medidas descritivas com o auxílio de calculadoras.

27

Estatística Descritiva: Tabelas

Tabelas: Distribuição de Frequência em Classes com Perda de Informação Observações

Notamos que cada um dos valores observados deve pertencer a uma e somente uma classe.

É usual que o limite inferior da primeira classe seja igual ao menor valor observado e que o maior valor pertença à última classe.

Quando o limite superior da última classe coincidir com o maior valor observado, é mais apropriado fechar este intervalo, contando o elemento nesta classe, do que abrir uma nova classe contendo apenas uma frequência absoluta.

Por outro lado, se o maior valor observado for inferior ao limite superior da classe, não há problemas, pois fixamos todas as classes com a mesma amplitude.

28

Estatística Descritiva: Gráficos

O que é um gráfico?

Gráfico é um recurso visual da estatística utilizado para representar um fenômeno.

Uma representação gráfica coloca em evidência as tendências, as ocorrências ocasionais, os valores mínimos e máximos e também as ordens de grandezas dos fenômenos que estão sendo observados.

Todo gráfico, em sua versão final, deve primar pela simplicidade, clareza e veracidade nas informações.

29

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Barras

O **gráfico de barras** é um gráfico formado por retângulos horizontais de larguras iguais, em que cada um deles representa a intensidade de uma modalidade ou atributo. O objetivo deste gráfico é comparar grandezas e é recomendável para variáveis cujas categorias tenham designações extensas.

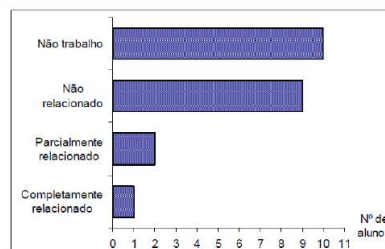


Figura 02 - Relação trabalho e curso dos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.
Fonte: DES/UEM.

30

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Colunas

O **gráfico de colunas** é o gráfico mais utilizado para representar variáveis qualitativas. Difere do gráfico de barras por serem seus retângulos dispostos verticalmente ao eixo das abscissas, sendo mais indicado quando as designações das categorias são breves.

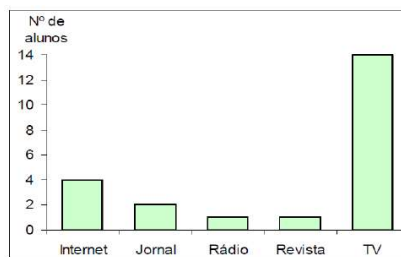


Figura 03 - Meios de informação utilizados pelos alunos da disciplina Inferência Estatística, curso de Estatística da UEM, 21/03/2005.
Fonte: DES/UEM.

31

Estatística Descritiva: Gráficos

Gráficos: Gráficos de Comparativos

Ao descrevermos, simultaneamente, duas ou mais categorias para uma variável, é conveniente fazer uso dos gráficos de barras ou colunas justapostas (ou sobrepostas), chamados de **gráficos comparativos**.

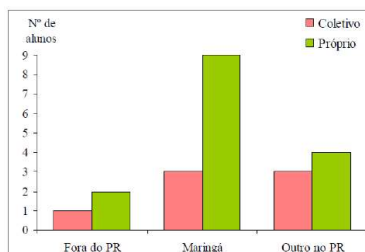


Figura 04 - Município de procedência segundo o tipo de transporte utilizado pelos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.
Fonte: DES/UEM.

32

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Setores

O **gráfico de setores** é um tipo de gráfico em que a variável em estudo é projetada num círculo, de raio arbitrário, dividido em setores com áreas proporcionais às **frequências** das suas categorias.

São indicados quando desejamos comparar cada valor da série com o total.

Recomenda-se o seu uso para o caso em que o número de categorias não é grande e não obedece a alguma ordem específica.

33

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Setores

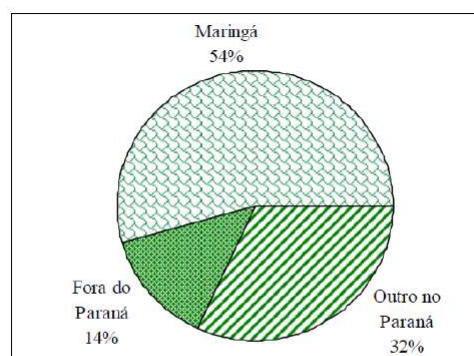


Figura 05 - Município de procedência dos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.

Fonte: DES/UEM.

34

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Setores

Cálculo do Ângulo em Gráfico de Setor

O procedimento para o cálculo do ângulo correspondente a cada categoria é feito por meio de simples proporções: 360 graus, que corresponde a um círculo completo, esta para o total de alunos entrevistados, 22, assim como x graus esta para o total de alunos que pertencem a categoria desejada.

Por exemplo, aos 54% de alunos que residem no município de Maringa correspondera um ângulo x resultante da expressão

$$\frac{360}{22} = \frac{x}{12},$$

cujo valor e aproximadamente 196 graus.

35

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Linhas

A aplicação de um [gráfico de linhas](#) é mais indicada para representações de séries temporais sendo, por tal razão, conhecidos também como gráficos de séries cronológicas.

Sua construção é feita colocando no eixo vertical (y) a mensuração da variável em estudo e na abscissa (x) as unidades da variável em uma ordem crescente.

Este tipo de gráfico permite representar séries longas, o que auxilia detectar suas flutuações tanto quanto analisar suas tendências. Também podemos representar várias séries em um mesmo gráfico.

36

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Linhas

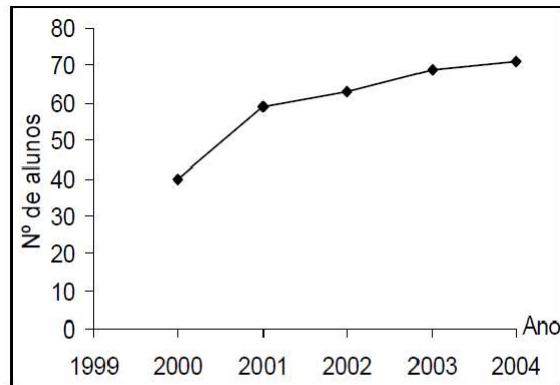


Figura 06 - Número de matrículas anuais na disciplina Probabilidade do curso de Estatística da UEM, 21/03/2005.

Fonte: DES/UEM.

37

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Bastões

Um [gráfico de bastões](#) é formado por segmentos de retas perpendiculares ao eixo horizontal (eixo da variável), cujo comprimento corresponde à frequência absoluta ou relativa de cada elemento da distribuição.

Suas coordenadas não podem ser unidas porque a leitura do gráfico deve tornar claro que não há continuidade entre os valores individuais assumidos pela variável em estudo.

38

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Bastões

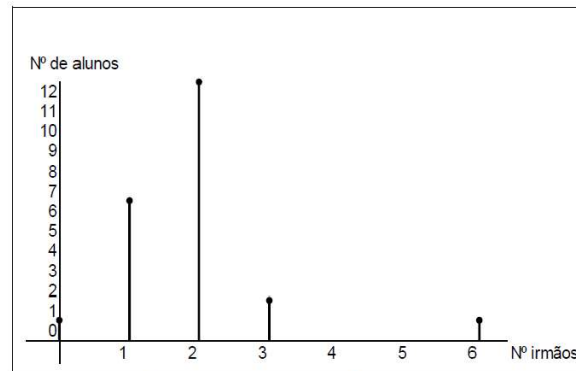


Figura 07 – Número de irmãos dos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.
Fonte: DES/UEM.

39

Estatística Descritiva: Gráficos

Gráficos: Gráficos de Frequência Acumulada

A Figura 08 mostra o [gráfico de frequência acumulada](#) de uma variável quantitativa discreta.

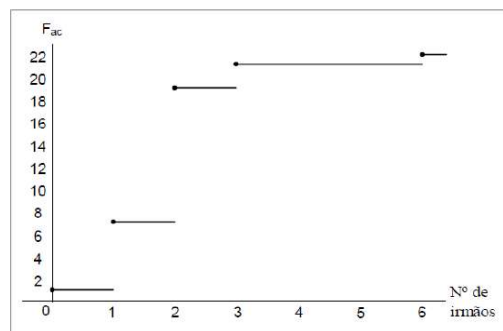


Figura 08 - Número acumulado de irmãos dos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.
Fonte: DES/UEM.

40

Estatística Descritiva: Gráficos

Gráficos: Histograma

O **histograma** é um gráfico de colunas justapostas que representa uma distribuição de frequência para dados contínuos ou uma variável discreta, quando esta apresentar muitos valores distintos.

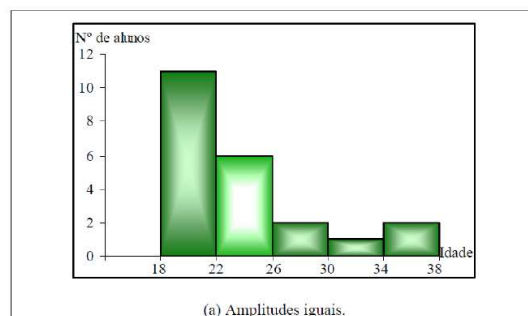
No eixo horizontal são dispostos os limites das classes segundo as quais os dados foram agrupados, enquanto que o eixo vertical corresponde às frequências absolutas ou relativas das mesmas.

41

Estatística Descritiva: Gráficos

Gráficos: Histograma

Quando os dados são distribuídos em classes de mesma amplitude, todas as colunas apresentam bases iguais com alturas variando em função das suas frequências absolutas ou relativas. Neste caso, a área de cada retângulo depende apenas da sua altura.

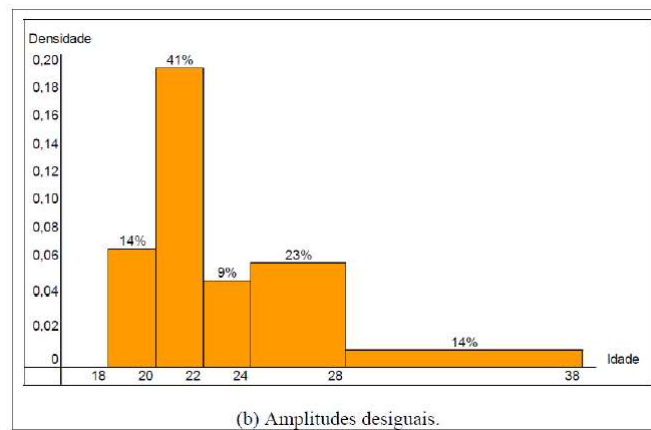


42

Estatística Descritiva: Gráficos

Gráficos: Histograma

No caso dos dados serem agrupados em classes de amplitudes diferentes, a área de cada coluna já não é mais proporcional à sua altura.



43

Estatística Descritiva: Gráficos

Gráficos: Histograma

Portanto, podemos dizer que, no primeiro caso, o eixo dos valores informa a frequência relativa de cada classe e, no segundo caso, tal procedimento perde todo significado e é necessário comparar as áreas para interpretar as informações que são expostas.

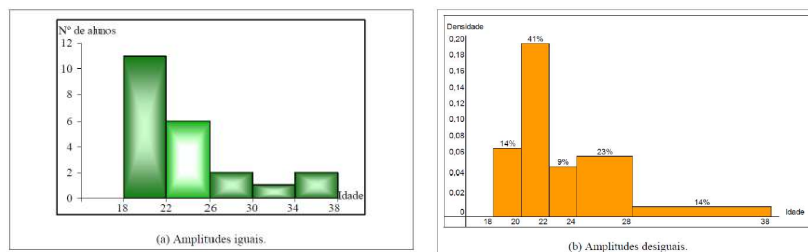


Figura 09 - Idade dos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.

Fonte: DES/UEM.

44

Estatística Descritiva: Gráficos

Gráficos: Polígono de Frequência

O **polígono de frequência** é um gráfico de linha cuja construção é feita unindo os pontos de coordenadas de abscissas correspondentes aos pontos médios de cada classe e as ordenadas, às frequências absolutas ou relativas dessas mesmas classes.

O **polígono de frequência** é um gráfico que deve ser **fechado no eixo das abscissas**. Então, para finalizar sua elaboração, devemos acrescentar à distribuição uma classe à esquerda e outra à direita, ambas com frequências zero.

Tal procedimento permite que a área sob a linha de frequências seja igual à área do histograma.

45

Estatística Descritiva: Gráficos

Gráficos: Polígono de Frequência

Uma das vantagens da aplicação de polígonos de frequências é que, por serem gráficos de linhas, permitem a comparação entre dois ou mais conjuntos de dados por meio da superposição dos mesmos.

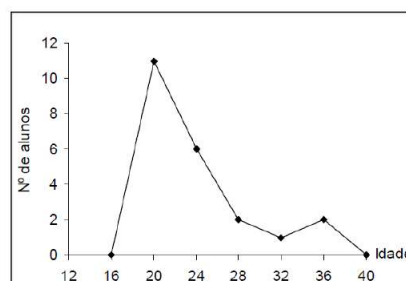


Figura 10 - Idade dos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.

Fonte: DES/UEM.

46

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Frequência Acumulada ou Ogiva

O **gráfico da frequência acumulada** ou **ogiva** é um gráfico que permite descrever dados quantitativos por meio da frequência acumulada.

A **ogiva** é um gráfico de linha que une os pontos cujas abscissas são os limites superiores das classes e as ordenadas são suas respectivas frequências acumuladas.

47

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Frequência Acumulada ou Ogiva

Convém observarmos que o ponto inicial desse gráfico é o limite inferior do primeiro intervalo, com frequência acumulada zero, pois não existe qualquer valor inferior a ele.

Quando os dados contidos em cada classe são **distribuídos uniformemente**, podemos estimar, a partir da ogiva, o número de elementos pertencentes a qualquer uma das classes que compõe a distribuição de frequência dos dados e a quantidade ou porcentagem de elementos que estão abaixo de certo valor pertencente ao conjunto de dados.

48

Estatística Descritiva: Gráficos

Gráficos: Gráfico de Frequência Acumulada ou Ogiva

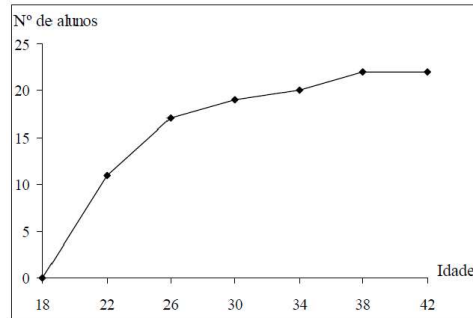


Figura 11 - Idade acumulada dos alunos da disciplina Inferência Estatística do curso de Estatística da UEM, 21/03/2005.

Analisando o gráfico acima, notamos que não existem alunos com idade inferior a 18 anos enquanto que abaixo de 34 anos existem vinte alunos.

49

Estatística Descritiva: Gráficos

Gráficos: Diagrama Ramo-e-Folhas

O **diagrama ramo-e-folhas** é um procedimento utilizado para armazenar os dados sem perda de informação.

É utilizado para termos uma ideia visual da distribuição dos dados.

Cada valor observado, x_i , da variável X , deve consistir de, no mínimo, dois dígitos e a variável pode ser tanto quantitativa discreta como contínua.

50

Estatística Descritiva: Gráficos

Gráficos: Diagrama Ramo-e-Folhas

Para construí-lo, dividimos cada número em duas partes. A primeira é denominada **ramo** e a segunda, **folhas**.

O ramo consistirá de um ou mais dígitos iniciais, se o valor da variável for um número inteiro, e do número inteiro, se o valor da variável for um número com decimais.

Nas folhas, colocamos os dígitos restantes, se o valor observado for número inteiro, ou os decimais, caso contrário.

51

Estatística Descritiva: Gráficos

Gráficos: Diagrama Ramo-e-Folhas

Ramo	Folha	Frequência
1	8 8 9	3
2	0 0 0 0 0 0 0 1 1 2 3 4 5 5 6 9	16
3	5 7	2

(a) Sem divisão de ramos.

Ramo	Folha	Frequência
1	8 8 9	3
2	0 0 0 0 0 0 0 1 1 2 3 4	12
2*	5 5 6 9	4
3	5 7	2

(b) Com divisão de ramos.

Observamos que o ramo correspondente ao dígito 2 tem muitas folhas. Neste caso, a opção e dividir este ramo em dois: as folhas de 0 a 4 pertencerão a uma linha e as folhas de 5 a 9 pertencerão a outra linha. Luciana

Figura 12 - Idade dos alunos da disciplina Inferência Estatística do curso de Estatística da Universidade Estadual de Maringá, 21/03/2005.

Fonte: DES/UEM.

52

Medidas Descritivas

Introdução

Uma outra maneira de resumirmos os dados de uma variável quantitativa, além de tabelas e gráficos, é apresentá-los na forma de valores numéricos, denominados **medidas descritivas**.

Estas medidas, se calculadas a partir de dados populacionais, são denominadas **parâmetros** e, se calculadas a partir de dados amostrais, são denominadas **estimadores** ou **estatísticas**.

As medidas descritivas auxiliam a análise do comportamento dos dados. Tais dados são provenientes de uma população ou de uma amostra, o que exige uma notação específica para cada caso.

53

Medidas Descritivas

Introdução

Classificamos as medidas descritivas como: **medidas de posição** (tendência central e separatrizes), **medidas de dispersão**, **medidas de assimetria** e **medidas de curtose**.

Notações

Medidas	Parâmetros	Estimadores
Número de elementos	N	n
Média	μ	\bar{X}
Variância	σ^2	S^2
Desvio padrão	σ	S

54

Medidas Descritivas

Medidas de Tendência Central

As **medidas de tendência central** são assim denominadas por indicarem um ponto em torno do qual se concentram os dados. Este ponto tende a ser o **centro** da distribuição dos dados.

A seguir, são definidas as principais medidas de tendencia central:

- Média
- Moda
- Mediana

55

Medidas Descritivas

Medidas de Tendência Central: Média Aritmética

A **media aritmética** (\bar{X}) é a soma de todos os valores observados da variável dividida pelo número total de observações.

Sob uma visão geométrica, a média de uma distribuição é o centro de gravidade e representa o ponto de equilíbrio de um conjunto de dados.

É a medida de tendencia central mais utilizada para representar um conjunto de dados.

56

Medidas Descritivas

Medidas de Tendência Central: Média Aritmética

Seja $X = \{x_1, \dots, x_n\}$ um conjunto de dados. A média aritmética é dada por:

$$\mu = \frac{\sum_{i=1}^N x_i}{N} \quad \text{ou} \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

para dados populacionais ou amostrais, respectivamente.

57

Medidas Descritivas

Medidas de Tendência Central: Média Aritmética

Caso os dados estejam apresentados segundo uma distribuição de frequências, temos:

$$\mu = \frac{\sum_{i=1}^k x_i F_i}{N} \quad \text{ou} \quad \bar{x} = \frac{\sum_{i=1}^k x_i F_i}{n}$$

Observe que, no caso de dados agrupados, a média é obtida a partir de uma ponderação, em que os pesos são as frequências absolutas de cada classe e x_i é o ponto médio da classe i .

58

Medidas Descritivas

Medidas de Tendência Central: Média Aritmética Propriedades

1. A média é um valor calculado facilmente e depende de todas as observações.
2. A média é única em um conjunto de dados e nem sempre tem existência real, ou seja, nem sempre é igual a um determinado valor observado.
3. A média é afetada por valores extremos observados.

59

Medidas Descritivas

Medidas de Tendência Central: Média Aritmética Propriedades

4. Por depender de todos os valores observados, qualquer modificação dos dados fará com que a média tenha a mesma alteração. Isto quer dizer que somando-se, subtraindo-se, multiplicando-se ou dividindo-se uma constante a cada valor observado, a média ficará acrescida, diminuída, multiplicada ou dividida por esse valor.
5. A soma da diferença de cada valor observado em relação à média é zero, ou seja, a soma dos desvios é zero

60

Medidas Descritivas

Medidas de Tendência Central: Moda

A **moda** (M_o) é o valor que apresenta a maior frequência da variável entre os valores observados.

Para o caso de valores individuais, a moda pode ser determinada imediatamente observando-se o rol ou a frequência absoluta dos dados.

Por outro lado, em se tratando de uma distribuição de frequência de valores agrupados em classes, primeiramente é necessário identificar classe modal, aquela que apresenta a maior frequência e, a seguir, a moda pode ser calculada.

61

Medidas Descritivas

Medidas de Tendência Central: Mediana

A **mediana** (M_d) é o valor que ocupa a posição central da série de observações de uma variável, em rol, dividindo o conjunto em duas partes iguais, ou seja, a quantidade de valores inferiores a mediana e igual à quantidade de valores superiores. Retomando o exemplo do número de filhos por famílias, verificamos que, para o caso de oito famílias, $n = 8$, a mediana é determinada como a seguir:

X	x_1	x_2	x_3	x_4		x_5	x_6	x_7	x_8
Valor observado	0	1	1	2	$\frac{x_4 + x_5}{2}$	2	2	3	4
← 4 observações →					$M_d = 2$	← 4 observações →			

62

Medidas Descritivas

Medidas de Tendência Central: Mediana

Quando acrescentamos ao grupo uma outra família com 10 filhos, o tamanho da amostra passa a ser $n=9$. Neste caso, a mediana é:

X	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9
Valor observado	0	1	1	1	2	2	3	4	10

\leftarrow 4 observações \rightarrow Md=2 \leftarrow 4 observações \rightarrow

63

Medidas Descritivas

Medidas Separatrizes

As **medidas separatrizes** são valores que ocupam posições no conjunto de dados, em rol, dividindo-o em partes iguais e podem ser:

- **Quartil:** Os quartis dividem o conjunto de dados em quatro partes iguais.
- **Decil:** Os decis dividem o conjunto de dados em dez partes iguais.
- **Percentil:** Os percentis dividem o conjunto de dados em cem partes iguais.

64

Medidas Descritivas

Medidas Separatrizes

- **Quartil:** Os quartis dividem o conjunto de dados em quatro partes iguais.

Estatística	Notação	Interpretação	Posição
1º quartil	Q_1	25% dos dados são valores menores ou iguais ao valor do primeiro quartil.	$p=0,25(n+1)$
2º quartil	$Q_2 = M_d$	50% dos dados são valores menores ou iguais ao valor do segundo quartil.	$p=0,50(n+1)$
3º quartil	Q_3	75% dos dados são valores menores ou iguais ao valor do terceiro quartil.	$p=0,75(n+1)$

65

Medidas Descritivas

Medidas Separatrizes

- **Decil:** Os decis dividem o conjunto de dados em dez partes iguais.

Estatística	Notação	Interpretação	Posição
1º decil	D_1	10% dos dados são valores menores ou iguais ao valor do primeiro decil.	$p=0,10(n+1)$
2º decil	D_2	20% dos dados são valores menores ou iguais ao valor do segundo decil.	$p=0,20(n+1)$
3º decil	D_3	30% dos dados são valores menores ou iguais ao valor do terceiro decil.	$p=0,30(n+1)$
4º decil	D_4	40% dos dados são valores menores ou iguais ao valor do primeiro decil.	$p=0,40(n+1)$
5º decil	$D_5 = Q_2 = M_d$	50% dos dados são valores menores ou iguais ao valor do segundo decil.	$p=0,50(n+1)$
6º decil	D_6	60% dos dados são valores menores ou iguais ao valor do terceiro decil.	$p=0,60(n+1)$
7º decil	D_7	70% dos dados são valores menores ou iguais ao valor do primeiro decil.	$p=0,70(n+1)$
8º decil	D_8	80% dos dados são valores menores ou iguais ao valor do segundo decil.	$p=0,80(n+1)$
9º decil	D_9	90% dos dados são valores menores ou iguais ao valor do terceiro decil.	$p=0,90(n+1)$

66

Medidas Descritivas

Medidas Separatrizes

- **Percentil:** Os percentis dividem o conjunto de dados em cem partes iguais.

Estatística	Notação	Interpretação	Posição
5º Percentil	P_5	5% dos dados são valores menores ou iguais ao valor do primeiro percentil.	$p=0,05(n+1)$
10º Percentil	P_{10}	10% dos dados são valores menores ou iguais ao valor do décimo percentil.	$p=0,10(n+1)$
25º Percentil	$P_{25}=Q_1$	25% dos dados são valores menores ou iguais ao valor do percentil cinquenta.	$p=0,25(n+1)$
50º Percentil	$P_{50}=D_5=Q_2=Md$	50% dos dados são valores menores ou iguais ao valor do primeiro percentil.	$p=0,50(n+1)$
75º Percentil	$P_{75}=Q_3$	75% dos dados são valores menores ou iguais ao valor do primeiro percentil. (Q_3)	$p=0,75(n+1)$
90º Percentil	P_{90}	90% dos dados são valores menores ou iguais ao valor do percentil noventa.	$p=0,90(n+1)$
95º Percentil	P_{95}	95% dos dados são valores menores ou iguais ao valor do percentil noventa e cinco.	$p=0,95(n+1)$

67

Medidas Descritivas

Medidas de Dispersão

Fenômenos que envolvem análises estatísticas caracterizam-se por suas semelhanças e variabilidades.

As **medidas de dispersão** auxiliam as **medidas de tendencia central** a descrever o conjunto de dados adequadamente.

As medidas de dispersão indicam se os dados estão, ou não, próximos uns dos outros.

68

Medidas Descritivas

Medidas de Dispersão

Desta forma, não há sentido em calcularmos a média de um conjunto em que não haja variação dos seus elementos. Existe ausência de dispersão e a medida de dispersão é igual a zero.

Por outro lado, aumentando a dispersão, o valor da média varia e, se a variação for muito grande, a média não será uma medida de tendência central representativa.

É necessário, portanto, ao menos **uma medida de tendência central** e **uma medida de dispersão** para descrever um conjunto de dados.

69

Medidas Descritivas

Medidas de Dispersão

As quatro medidas de dispersão que serão definidas a seguir são:

- Amplitude Total,
- Amplitude Interquartilíca,
- Variância,
- Desvio Padrão.

Com exceção da primeira, todas têm como ponto de referência a **média**.

70

Medidas Descritivas

Medidas de Dispersão: Amplitude Total

A **amplitude total** de um conjunto de dados é a diferença entre o maior e o menor valor observado.

$$AT = x_{max} - x_{min}$$

Esta medida de dispersão não leva em consideração os valores intermediários, perdendo a informação de como os dados estão distribuídos e/ou concentrados.

71

Medidas Descritivas

Medidas de Dispersão: Amplitude Interquartílica

A **amplitude interquartílica** é a diferença entre o terceiro e o primeiro quartil.

Esta medida é mais estável que a amplitude total por não considerar os valores mais extremos.

Esta medida abrange 50% dos dados e é útil para detectar valores discrepantes.

$$d_q = Q_3 - Q_1$$

72

Medidas Descritivas

Medidas de Dispersão: Desvios

A diferença entre cada valor observado e a média é denominado desvio e é dado por $(x_i - \mu)$ se o conjunto de dados é populacional e por $(x_i - \bar{x})$ se os dados são amostrais.

Ao somar todos os desvios, ou seja, ao somar todas as diferenças de cada valor observado em relação a média, o resultado é igual a zero (propriedade 5 da média).

Isto significa que esta medida não mede a variabilidade dos dados.

73

Medidas Descritivas

Medidas de Dispersão: Desvios Desvio Médio

Para resolvermos este problema, podemos considerar as diferenças em módulo. A média destas diferenças, em módulo, é denominada desvio médio:

$$d_m = \frac{\sum_{i=1}^N |x_i - \mu|}{N} \text{ ou } d_m = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{N}$$

para dados populacionais ou amostrais, respectivamente. Caso os dados estejam apresentados segundo uma distribuição de frequência, temos:

$$d_m = \frac{\sum_{i=1}^N |x_i - \mu| F_i}{N} \text{ ou } d_m = \frac{\sum_{i=1}^n |x_i - \bar{x}| F_i}{N}$$

74

Medidas Descritivas

Medidas de Dispersão: Desvios Variância

Não há nada conceitualmente errado em considerarmos o desvio médio, mas esta medida não tem certas propriedades importantes e não é muito utilizada.

O mais comum é considerarmos o quadrado dos desvios em relação à média e, então, calcularmos a medida. Obtemos, assim, a **variância**, que é definida por:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2}{N} \text{ ou } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

se os dados são populacionais ou amostrais, respectivamente.

75

Medidas Descritivas

Medidas de Dispersão: Desvios Variância

Caso os dados estejam apresentados segundo uma distribuição de frequência, temos:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \mu)^2 F_i}{N} \text{ ou } s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 F_i}{n-1}$$

76

Medidas Descritivas

Medidas de Dispersão: Desvios Desvio Padrão

Entretanto, ao calcularmos a variância, observamos que o resultado será dado em unidades quadráticas, o que dificulta a sua interpretação. O problema é resolvido extraindo-se a raiz quadrada da variância, definindo-se, assim, o **desvio padrão**:

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2}{N}} \text{ ou } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

se os dados são populacionais ou amostrais.

77

Medidas Descritivas

Medidas de Dispersão: Desvios Desvio Padrão

Se estiverem em distribuição de frequências, o desvio padrão é definido como

$$\sigma = \sqrt{\frac{\sum_{i=1}^N (x_i - \mu)^2 F_i}{N}} \text{ ou } s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2 F_i}{n-1}}$$

Geralmente, o desvio padrão é maior ou igual ao desvio médio, pois o cálculo do desvio padrão considera cada desvio elevado ao quadrado, aumentando desproporcionalmente o peso dos valores extremos.

78

Medidas Descritivas

Medidas de Dispersão: Desvios Coeficiente de Variação

O **coeficiente de variação** é uma medida de dispersão relativa definida como a razão entre o desvio padrão e a média:

$$CV = \frac{\sigma}{\mu} \times 100 \text{ ou } CV = \frac{s}{\bar{x}} \times 100,$$

se os dados são populacionais ou amostrais.

79

Medidas Descritivas

Medidas de Dispersão: Desvios Coeficiente de Variação

A partir do coeficiente de variação podemos avaliar a homogeneidade do conjunto de dados e, consequentemente, se a média é uma boa medida para representar estes dados.

O coeficiente de variação também pode ser utilizado para comparar conjuntos com unidades de medidas distintas.

Uma desvantagem do coeficiente de variação é que ele deixa de ser útil quando a média está próxima de zero. Uma média muito próxima de zero pode in acionar o coeficiente de variação.

80

Medidas Descritivas

Medidas de Dispersão: Desvios Coeficiente de Variação

Um coeficiente de variação superior a 50% sugere alta dispersão, o que indica heterogeneidade dos dados. Quanto maior for este valor, menos representativa será a média.

Neste caso, optamos pela mediana ou moda, não existindo uma regra prática para a escolha de uma destas medidas. O pesquisador, com sua experiência, é que deverá decidir por uma ou outra.

Por outro lado, quanto mais próximo de zero, mais homogêneo e o conjunto de dados e mais representativa será a sua média.

81

Medidas Descritivas

Medidas de Assimetria

A medida de assimetria é um indicador da forma da distribuição dos dados. Ao construir uma distribuição de frequências e/ou um histograma, estamos buscando, também, identificar visualmente a forma da distribuição dos dados, fornecida pelo coeficiente de assimetria de Pearson (A_s), definido como:

$$A_s = \frac{\mu - M_o}{\sigma} \text{ ou } A_s = \frac{\bar{x} - M_o}{s};$$

para dados populacionais e amostrais, respectivamente.

82

Medidas Descritivas

Medidas de Curtose

A medida de curtose representa o grau de achatamento de uma distribuição e é um indicador da forma desta distribuição. É definida como:

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

A curtose é mais uma medida com a finalidade de complementar a caracterização da dispersão em uma distribuição.

Esta medida quantifica a concentração ou dispersão dos valores de um conjunto de dados em relação às medidas de tendência central em uma distribuição de frequências.

83

Medidas Descritivas

Medidas de Curtose

Classificação das Distribuições

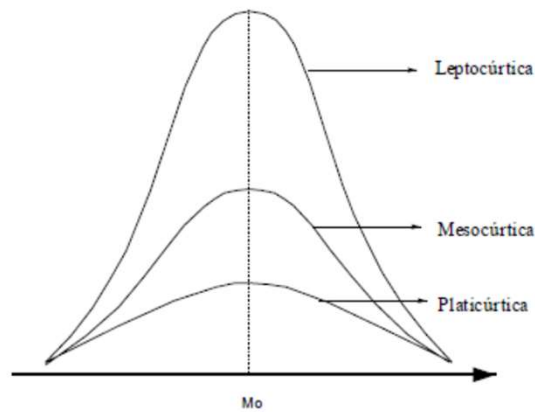
Uma distribuição é classificada quanto ao grau de achatamento como:

- Leptocúrtica: quando a distribuição apresenta uma curva de frequência bastante fechada, com os dados fortemente concentrados em torno de seu centro, $K < 0$; 263.
- Mesocúrtica: quando os dados estão razoavelmente concentrados em torno de seu centro, $K = 0$; 263.
- Platicúrtica: quando a distribuição apresenta uma curva de frequência mais aberta, com os dados fracamente concentrados em torno de seu centro, $K > 0$; 263.

84

Medidas Descritivas

Medidas de Curtose Classificação das Distribuições



85

Medidas Descritivas

Gráficos: Box Plot

O gráfico Box Plot (ou desenho esquemático) é uma análise gráfica que utiliza cinco medidas estatísticas:

- Mínimo
- Máximo
- Mediana
- Primeiro Quartil
- Terceiro Quartil

86

Medidas Descritivas

Gráficos: Box Plot

Este conjunto de medidas oferece a ideia da posição, dispersão, assimetria, caudas e dados discrepantes.

A posição central é dada pela mediana e a dispersão pelo desvio interquartilico, $dq = Q_3 - Q_2$.

As posições relativas de Q_1 , Q_2 e Q_3 dão uma noção da assimetria da distribuição.

Os comprimentos das caudas são dados pelas linhas que vão do retângulo aos valores atípicos.

87

Medidas Descritivas

Gráficos: Box Plot Outlier

Um **outlier**, ou valor atípico, é um valor que se localiza distante de quase todos os outros pontos da distribuição.

A distância a partir da qual consideramos um valor como atípico é aquela que supera $1,5dq$.

De maneira geral, são considerados outliers todos os valores inferiores a $L_i = Q_1 - 1,5dq$ ou os superiores a $L_s = Q_3 + 1,5dq$.

88

Medidas Descritivas

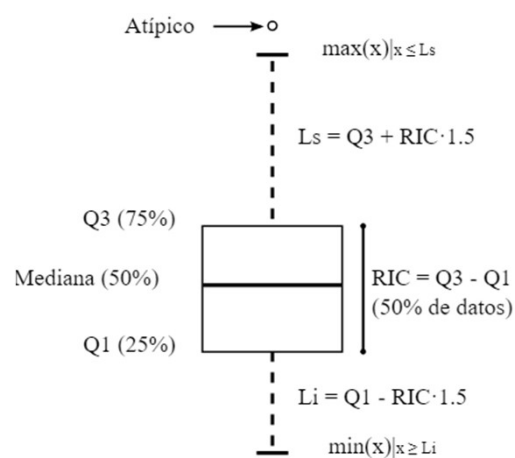
Gráficos: Box Plot

O gráfico Box Plot pode ser utilizado para fazer comparações entre várias distribuições. Essa comparação é feita através de vários desenhos esquemáticos numa mesma figura.

89

Medidas Descritivas

Gráficos: Box Plot



90