
Medidas Descritivas

META:

É necessário usar técnicas de estatística descritiva sobre os dados de forma que possamos apresentá-los de forma mais simples e compacta, tornando seu estudo mais eficiente. Assim podemos lançar mão, inicialmente, do uso de Descrição Tabular e de Gráficos descritivos. Até este ponto estudamos as distribuições de frequências que nos permitiu, de modo geral, descrever os grupos de valores que uma variável pode assumir. Dessa forma, podemos localizar a maior concentração de valores de uma dada distribuição de uma variável, verificando em que parte do intervalo de abrangência dos dados, esta concentração está localizada. Entretanto, para se ressaltar as tendências características de uma dada distribuição, isoladamente, ou em confronto com outras, surge a necessidade de introduzir conceitos que se expressem de forma numérica essas tendências, e que nos permitam traduzi-las.

OBJETIVOS:

Ao fim da aula os alunos deverão ser capazes de construir de um gráfico usando qualquer software Estatístico. Este é um recurso usado para visualização dos dados de forma mais simplificada. As Medidas descritivas são métodos que objetiva, através de medidas numéricas únicas o resumo dos dados ou algumas de suas características em um único valor. Sendo estas medidas usadas

para a comparação entre diferentes amostras ou populações. São classificadas em Medidas de Tendência Central ou de Posição e Medidas de Dispersão ou de Variabilidade

PRÉ-REQUISITOS

Ter em mãos, esquadros, compasso, calculadora, papel milimetrado e quadriculado.

2.1 Introdução

Caros alunos nesta segunda aula do nosso curso estudaremos os gráficos mais usados na Estatística Descritiva: histograma; polígono de frequência; ogiva de Galton; e setorgrama. Para outras aplicações recomendo uma busca nos Softwre Estatísticos. Recomendo também o usos de uma calculadora científica para uso das Medidas Descritivas. A partir do uso das Medidas Descritivas começaremos o cálculo de alguns parâmetros como: média aritmética, moda e mediana, dos quais a média é o mais importante além de ser o de maior utilização. Após os cálculos iniciais o aluno irá identificar o significado de Dispersão. Com esta medida vai poder mensurar o afastamento médio entre todos os valores de uma série e sua medida de tendência central. Dispondo desta informação vai poder classificar qualquer conjunto de dados quanto à homogeneidade e heterogeneidade e trabalhar de forma precisa esta característica nas populações ou amostras que venha a investigar(SAMUEL,2009).

OBS 2.1. Caro aluno para um melhor entendimento dos gráficos é necessário buscar o recurso dos softweres Estatístico, como também procurar exemplos em revistas e jornais. Aqui definiremos apenas os conceitos básicos de gráfico pois, entendo que construir gráficos manualmente é um desperdício uma vez que, existem computadores que fazem este tipo de trabalho e com isto o aluno ganhará tempo e melhorará seu apredizado.

2.2 Histograma

O histograma é um gráfico formado por um conjunto de retângulos justapostos, de forma que a área de cada retângulo seja proporcional à frequência da classe que ele representa. Utilizado para representar as frequências, absolutas (f_i) ou relativas (fr_i), em relação à sua classe (ou intervalo de classe), é construído da seguinte forma:

1. No eixo X (abscissas), marcamos, em escala, os intervalos de classe.
2. No eixo Y (ordenadas), marcamos as frequências de classe (f_i) ou (fr_i).
3. Fazemos a correspondência entre cada intervalo no eixo dos X (classes) e um valor no eixo vertical (f_i), neste gráfico cada intervalo de classe será a base de um retângulo cuja altura é a respectiva frequência do intervalo. Ao final o gráfico forma um desenho que lembra um conjunto de colunas paralelas (retangulares) ou um conjunto de prédios.

2.3 Polígono de Frequência

Utilizado para indicar o ponto médio (Pm) ou qualquer outro ponto dentro de cada intervalo, como representante da classe com suas respectivas frequências, é construído sobre o histograma, o Polígono de Frequência. Os passos para sua a sua construção é apresentado a seguir:

- No eixo X (abscissas), colocamos o ponto médio de cada intervalo de classe.

- No eixo Y (ordenadas), permanecem as frequências absolutas de classe (f_i).
- Ligamos os pontos médios (Pm) das bases superiores dos retângulos do histograma por segmentos de reta.
- Para completar o polígono, acrescentamos um ponto médio com frequência zero em cada uma das extremidades da escala horizontal.

2.4 Setograma

Também conhecido como gráfico de pizza, é utilizado para representar os valores relativos em porcentagem (%).

1. Fazemos um círculo.
2. Cada Setor/Fatia é regido pela fórmula :

$$\text{Setor}^\circ = \frac{360^\circ \cdot f_i}{\text{Total}} \quad (2.14)$$

3. No círculo, distribuimos os valores das frequências percentuais.

2.5 Média Aritmética

Medida de tendência central que representa o ponto de concentração dos valores de um conjunto de dados ou uma sequência numérica. Esta pode ser simples ou ponderada. Em que a média aritmética simples de um conjunto numérico, é igual ao quociente entre a soma de todos os valores do conjunto e o número total de valores. Já a média ponderada é igual ao quociente entre a soma

de todos os valores, multiplicados pelo seus respectivos e a soma dos pesos. Para seu cálculo devemos considerar alguns se os dados estão ou não agrupados.

- Para os dados não agrupados (dados brutos)

$$\bar{X} = \sum_{i=1}^n \frac{x_i}{n} = \frac{(x_1 + x_2 + x_3 + x_4 + \dots + x_n)}{n} \quad (2.15)$$

em que x_i é o valor da i -ésima observação da amostra.

- Dados agrupados em distribuição de frequência simples, sem intervalo de classes. Consiste em uma média aritmética ponderada, cujos pesos são as frequências f_i definida por:

$$\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} \quad (2.16)$$

- Dados agrupados com intervalos de classe. Neste caso usamos os pontos médios das classes (Pm_i) como valor da frequência de cada classe. Definimos (Pm_i) como

$$Pm_i = \frac{L_i - l_i}{2} \quad (2.17)$$

em que L_i representa a o limite superior da classe e l_i representa a o limite inferior da classe. E para a média temos

$$\bar{X} = \frac{\sum_{i=1}^n f_i Pm_i}{\sum_{i=1}^n f_i} \quad (2.18)$$

2.6 Mediana

A mediana corresponde ao valor que divide um conjunto de dados ordenados de tal forma que pelo menos a metade ou 50%

dos itens sejam iguais ou menores são menores do que este, e que haja pelo menos outra metade ou 50% maiores. Sua notação é Md [?].

Elemento Mediano Como a mediana divide a distribuição em duas partes, podemos concluir que esta mediana apresenta um número de ordem. Assim a mediana é um valor que ocupa uma determinada ordem ou posição na série de dados. O número que indica a ordem em que se encontra o valor correspondente à mediana é denominado elemento mediano[?]. Notação E_{Md}

Para calculo da mediana devemos considerar se os dados estão tabulados mas não agrupados, tabulados e agrupados em classe ou não tabulados. E assim, para cada caso o cálculo se diferencia.

Dados não Tabulados

Estando os dados não tabulados (dados brutos), para obter a mediana, o procedimento é o seguinte. Primeiro colocamos a sequência numérica em ordem crescente ou decrescente. Em seguida localizamos o(s) valor(es) mediano(s) de acordo com a quantidade. Se o conjunto de dados for par ou impar devemos adotar um dos procedimentos a seguir:

- Se n é ímpar, o rol admite apenas um termo central que ocupa a posição $E_{Md} = \frac{n+1}{2}$.
- Se n é par o rol admite dois termos centrais, a saber os elementos de ordem $E_{Md1} = \frac{n}{2}$ e $E_{Md2} = \frac{n+1}{2}$. Então o valor da mediana será a média destes termos. Por exemplo sendo

Dados Tabulados não Agrupados em Classes

Quando os dados já estão tabulados, o procedimento é praticamente idêntico ao aplicado a dados não tabulados. Primeiro devemos verificar se o número de observações é ímpar ou par. Em seguida verificamos qual das frequências F_i possuem $n/2$ ou mais de 50% das observações do conjunto de dados.

Cálculo da mediana apresentados a seguir:

Valores	Frequências
x_j	f_i
2	5
3	10
4	15
5	12
6	5
7	3
	$n = 50$

Dados Tabulados Agrupados em Classes Para seu cálculo devemos seguir algumas etapas.

- Localizar a classe da mediana que corresponde à classe cujo valor da Frequência Acumulada F_i é imediatamente superior a $n/2$;
- Identificada a classe da mediana passamos ao cálculo do valor da mediana que consiste na equação 2.19. Calcula-se a ordem:

$$Md = l + c \times \frac{E_{Md} - F_{ant}}{f_{Md}} \quad (2.19)$$

em que

E_{Md} posição do valor da mediana no rol de dados, definido por

$$E_{Md} = \frac{n}{2};$$

F_{ant} frequência acumulada até a casa anterior;

f_{Md} frequência simples da classe da mediana;

l limite inferior da classe da mediana;

c amplitude da classe da mediana.

O cálculo da mediana pode também pode ser feito a partir das frequências relativas. Neste caso temos

$$Md = l + c \times \frac{0,5 - Fr_{ant}}{fr_{Md}} \quad (2.20)$$

em que

Fr_{ant} Frequência Acumulada Relativa até a classe anterior à mediana;

fr_{Md} Frequência Simples Relativa da classe da mediana;

Em algumas situações gostaríamos de verificas qual, ou quais valores são mais freqüentes no conjunto de dados. Este desejo do pesquisador arremete a seguinte definição.

2.7 Moda

Valor mais frequente(s) no conjunto de dados quando sua frequência comparada com os demais valores pertencentes ao mesmo conjunto de dados. Notação é Mo .

OBS 2.2. Um conjunto pode ter mais de uma moda.

Como foi observado para a mediana o cálculo da moda é diferente para os dados que estejam tabulados mas não agrupados ou tabulados e agrupados em classe ou não tabulados.

- NDados não Tabulados e não Agrupados em Classes Pela definição basta buscar o valor predominante, de maior frequência, no conjunto de dados;
- NDados Tabulados e Agrupados em Classes Dada a tabela ao lado devemos seguir alguns passos.
 1. : Identifica-se a classe modal (aquela que possui maior frequência).
 2. neste ponto temos dois métodos
 - (a) o método mais simples, chamado de média bruta, consiste em tomar o ponto médio da classe modal como o valor de Mo ;
 - (b) o método mais elaborado consiste em aplicação da formula de Czuber definida por

$$No = l + c \times \frac{\Delta_1}{\Delta_1 + \Delta_2} \quad (2.21)$$

em que

l é o limite inferior da classe modal;

c é a amplitude da classe modal;

Δ_1 que definido por $\Delta_1 = f_{mo} - f_{ant}$;

Δ_2 que definido por $\Delta_2 = f_{mo} - f_{pot}$.

f_{mo} a frequência da classe modal, isto é, a classe com maior frequência;

f_{ant} a frequência da classe anterior a classe modal;

f_{pos} a frequência da classe posterior a classe modal.

2.8 Desvio médio

O Desvio médio ou média dos desvios é igual a média aritmética dos valores absolutos dos desvios tomados em relação a média sendo escrito da seguinte forma:

- para dados não tabulados

$$D_m = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}. \quad (2.22)$$

- para dados tabulados não agrupados em classes

$$D_m = \frac{\sum_{j=1}^k |x_j - \bar{x}| f_j}{\sum_{j=1}^k f_j = n}. \quad (2.23)$$

- para dados tabulados agrupados em classes

$$D_m = \frac{\sum_{j=1}^k |Pm_j - \bar{x}| f_j}{\sum_{j=1}^k f_j = n}. \quad (2.24)$$

2.9 Variância

A Variância é igual a média aritmética dos valores dos desvios tomados em relação a média elevados ao quadrado, sendo escrito da seguinte forma:

- para dados não tabulados considerando toda a população.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}. \quad (2.25)$$

- para dados não tabulados

OBS 2.3. Variância considerando uma amostra.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}. \quad (2.26)$$

- para dados tabulados não agrupados em classes considerando toda a população.

$$S^2 = \frac{\sum_{j=1}^k (x_j - \bar{x})^2 f_j}{\sum_{j=1}^k f_j = n}. \quad (2.27)$$

OBS 2.4. Variância considerando uma amostra.

$$S^2 = \frac{\sum_{j=1}^k (x_j - \bar{x})^2 f_j}{\sum_{j=1}^k f_j = n - 1}. \quad (2.28)$$

- para dados tabulados agrupados em classes considerando toda a população.

$$S^2 = \frac{\sum_{j=1}^k (Pm_j - \bar{x}) f_j}{\sum_{j=1}^k f_j = n}. \quad (2.29)$$

OBS 2.5. Variância considerando uma amostra.

$$S^2 = \frac{\sum_{j=1}^k (Pm_j - \bar{x}) f_j}{\sum_{j=1}^k f_j = n - 1}. \quad (2.30)$$

2.10 Dispersão Relativa

Mede a relação entre a dispersão absoluta e o valor médio da série, com resultado expresso em porcentagens.

Tem grande aplicação, visto que possibilita comparar dispersão de universos diferentes (em unidade de medida e/ou valor médio), já que seus resultados são expressos em porcentagens devido à relação entre as medidas de dispersão absoluta e de tendência central. São medidas usadas para análises comparativas.

Coefficiente de variação de PEARSON - é a relação percentual entre o desvio padrão e a média aritmética.

$$CV = \frac{S}{\bar{X}} \times 100. \quad (2.31)$$

2.11 Exemplos

Nada mais natural que ilustrar um novo conceito com exemplos e, vamos aqui fazer exatamente isto.

exemplo1. Determinar a média aritmética simples dos valores: 3, 7, 8, 10 e 11

SOLUÇÃO:

A média é $\bar{X} = (3, 7, 8, 10, 11)/5 = 7,8$

Dada a amostra: 2, 5, 5, 5, 5, 6, 6, 6, 8, 8 calcule a média.

Baseado na amostra temos os cálculos resumido na Tabela a seguir.

x_i	f_i	$x_i f_i$
2	1	2
5	4	20
6	3	18
8	2	16
Total	10	56

logo,

SOLUÇÃO:

$$\bar{X} = \frac{\sum_{i=1}^n x_i f_i}{\sum_{i=1}^n f_i} = \frac{56}{10}$$

Calcule a média baseando-se nos dados da Tabela de frequência.

SOLUÇÃO:

lodo

$$\bar{X} = \frac{\sum_{i=1}^n x_i Pm_i}{\sum_{i=1}^n f_i} = \frac{43,79}{25} = 1,7516 = 1,75.$$

exemplo2. De posse do conjunto de dados 5, 13, 10, 2, 18, 15, 6, 16, 9 calcule a mediana para os seguintes casos.

SOLUÇÃO:

Tabela 2.1: Tabela de frequência das alturas.

Alturas	f_i	Pm_i	$f_i \cdot Pm_i$
1,70 † 1,72	3,00	1,71	5,13
1,72 † 1,74	4,00	1,73	6,92
1,74 † 1,76	10,00	1,75	17,50
1,76 † 1,78	5,00	1,77	8,85
1,78 † 1,80	2,00	1,79	3,58
1,80 † 1,81	1,00	1,81	1,81
Total	25,00		43,79

a) considerando $n = 9$

Colocando os valores em ordem crescente temos $\{2, 5, 6, 9, 10, 13, 15, 16, 18\}$,
temos que $Md = 10$

b) considerando $n = 8$, temos que $Md = \frac{9+10}{2} = 9,5$

exemplo3. Tendo em mãos os conjuntos de dados $A = \{7, 8, 8, 9, 10, 10, 10, 12, 15\}$,
 $B = \{3, 5, 8, 10, 12, 13\}$, $C = \{2, 2, 5, 5, 8, 9\}$

SOLUÇÃO:

1. para A o elemento de maior frequência é o 10, portanto $Mo = 10$ (unimodal)
2. para B todos os elementos da série apresentam a mesma frequência, logo a série não possui moda (amodal).
3. para C temos $Mo_1 = 2$ e $Mo_2 = 5$ (bimodal)

exemplo4. Determinar a variância das amostras constituída dos seguintes elementos, 7; 10; 12; 15; 16; 18 e 20.

SOLUÇÃO:

Passo 1

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{49 + 16 + 4 + 1 + 4 + 46 + 36}{6} = 16. \quad (2.32)$$

$$S = \sqrt{16} = 4. \quad (2.33)$$

exemplo5. Determinar a dispersão relativa da amostra constituída dos seguintes elementos, 7; 10; 12; 15; 16; 18 e 20.

SOLUÇÃO:

Passo 1

$$\bar{X} = \sum_{i=1}^n \frac{x_i}{n} = \frac{(7 + 10 + 12 + 15 + 16 + 18 + 20)}{7}. \quad (2.34)$$

$$\bar{X} = \frac{98}{7} = 14. \quad (2.35)$$

Passo 2

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} = \frac{49 + 16 + 4 + 1 + 4 + 46 + 36}{6} = 16. \quad (2.36)$$

$$S = \sqrt{16} = 4. \quad (2.37)$$

Passo 3

$$CV = \frac{S}{\bar{X}} \times 100 = \frac{4}{14} \times 100 = 28,57\%. \quad (2.38)$$

OBS 2.6. Uma fonte de confusão comum é que o termo variância da amostra com a notação S^2 é quando calculamos a variância usando $n - 1$ em vez de n . Quando usamos $n - 1$ estamos fazendo uma correção por usarmos a média da amostra \bar{x} como uma estimativa da média populacional μ , o que não conhecemos. Na prática, porém, para grandes amostras o uso de n ou $n - 1$ provocam individualmente uma pequena diferença.

Observando o cálculo da variância, podemos notar que dados que possuem uma unidade² terá sua unidade elevada ao quadrado. Então para corrigir este problema, no momento de uma comparação feita pelo pesquisador, por exemplo, com um conjunto de dados da mesma natureza, podemos usar o Desvio Padrão.

O Desvio padrão é apenas a raiz quadrada da Variância com a notação $S = \sqrt{S^2}$.

²por exemplo, o peso, cuja unidade pode ser metro (m) qualquer, como por exemplo, centímetro (cm)

2.12 Conclusão

Na aula de hoje, vimos um recurso usado para visualização dos dados de forma mais simplificada que são os gráficos são. histograma; polígono de frequência; ogiva de Galton; e setorgrama. Vimos que as medidas de posição podem-se apresentar de várias formas dependendo daquilo que se pretende conhecer. A respeito da variável de interesse. São muitas vezes calculadas, no início da análise de um conjunto de dados. Após todos os cálculos de medidas que exprimem posição dos dados, podemos nos questionar, pensando qual a abrangência dos meus dados? Será que meus dados são pouco e muitos espalhados em relação a alguma referência? Assim a partir deste momento apresentaremos medidas estatísticas que nos deem respostas a estas perguntas. Estas medidas são chamadas de medidas de dispersão. Desta classe de medidas de variabilidade, as mais comuns são a Amplitude, O desvio Médio, a Variância e o Desvio Padrão, e o Coeficiente de Variação.



RESUMO

Aprendemos a construir gráficos, entretanto, é necessário o uso dos softwares estatísticos na construção de um gráfico. Vimos que as medidas de posição resumem os dados ou algumas de suas características em um único valor. Vimos também as Medidas de Dispersão, dando ênfase aos de uso mais geral como: a variância, o desvio padrão e o coeficiente de variação. Além do cálculo destes parâmetros é importante saber interpretar e entender suas finalidades e quando podem ser utilizados. Lembre-se que você partiu

de informações oriundas de uma pesquisa ou de um banco de dados, organizou e pode resumir tudo em tabelas de Frequências, Gráficos e Medidas Descritivas(SAMUEL, 2009).

PRÓXIMA AULA

Na próxima aula estudaremos os conceitos básicos de Probabilidade e as definições de Esperança e Variância de uma Variável Aleatória.

ATIVIDADES

OBS 2.7. Todas as atividades serão respondidas na aula presencial.

ATIV. 2.1. Dez medidas do diâmetro de um cilindro foram anotadas por um cientista como: 3,88; 4,09; 3,92; 3,97; 4,02; 3,95; 4,03; 3,92; 3,98 e 4,06 polegadas. Determinar a média aritmética.

Comentário: Volte ao texto e reveja com calma e atenção os exemplos acima, eles lhe servirão de guia.

ATIV. 2.2. Achar a mediana dos seguintes valores: 46, 47, 49, 51, 54, 53, 53, 57, 58, 58, 59, 62, 65, 67, 71, 71, 73, 76 e 79.

Comentário: Volte ao texto e reveja com calma e atenção os exemplos acima, eles lhe servirão de guia.

ATIV. 2.3. Um hospital calculou a média de gastos semestral e o desvio padrão da conta com pacientes internados na UTI hospitalar para cada uma das suas filiais. Na filial A, a média foi de 16000 mil com desvio-padrão de 2000 mil reais. Na filial B a conta foi de



15000 mil com desvio-padrão de 5000 mil reais. explique o que se pode inferir a partir desses valores.

Comentário: Volte ao texto e reveja com calma e atenção os exemplos acima, eles lhe servirão de guia.

$$CV_A = \frac{S}{\bar{X}} \times 100 = \frac{2000}{16000} \times 100 = 12,5. \quad (2.39)$$

$$CV_B = \frac{S}{\bar{X}} \times 100 = \frac{5000}{15000} \times 100 = 33,3. \quad (2.40)$$

ATIV. 2.4. O conjunto, a seguir, representa as notas obtidas por 15 alunos em uma prova de Inglês. 1, 2, 3, 8, 5, 7, 6, 9, 4, 6, 2, 10, 3, 5, 3. A mediana desse conjunto será:

Comentário: Volte ao texto e reveja com calma e atenção os exemplos acima, eles lhe servirão de guia.

a)nota 5,0 b)5,0 alunos c)9,0 alunos d)nota 9,0

ATIV. 2.5. Trinta e seis estudantes foram submetidos a um exame de biologia, obtendo as seguintes notas. Determine a média, a mediana e o desvio padrão. 93 98 72 84 88 90 78 80 89 94 95 77 81 75 72 76 70 76 83 99 73 83 87 91 83 92 90 92 77 86 86 74 87 81 98 94

Comentário: Volte ao texto e reveja com calma e atenção os exemplos acima, eles lhe servirão de guia.

ATIV. 2.6. As notas finais de um estudante em matemática, estatística, português e história são respectivamente 82, 86, 91e 71, atribuindo-se a essas disciplinas, respectivamente, os pesos 3, 5, 3, 1 a média aritmética será.

Comentário: Volte ao texto e reveja com calma e atenção os exemplos acima, eles lhe servirão de guia.

AUTO-AVALIAÇÃO

Sou capaz de entender a diferença o que é Média, Moda e mediana ?

Sou capaz de elaborar e diferenciar as Medidas de Posição das Medidas de Variabilidade?

LEITURA COMPLEMENTAR

CARVALHO, Sérgio, Estatística Básica: Série Impetus Provas e Concursos, Rio de Janeiro, 2^a edição, 2006.

TOLEDO, Geraldo Luciano, OVALLE, Ivo Izidoro, estatística básica. 2^a edição, editora Atlas, 1995.

FRANCISCO ESTEVAM MARTINS DE OLIVEIRA. Estatística e Probabilidade. Editora Atlas.

SAMUEL DE OLIVEIRA RIBEIRO, Métodos Quantitativos em Biologia. LUIZ A. C. GÓES. Estatística I e II. Editora Saraiva.

FRANCISCA DÍAZ; FRANCISCO JAVIER LOPES. Bioestatística. Editora Thomson.