Apostila

De

Estatística

Professores: Wanderley Akira Shiguti

Valéria da S. C. Shiguti

Brasília 2006

INTRODUÇÃO

1.1. PANORAMA HISTÓRICO

- Toda Ciência tem suas raízes na história do homem;
- A Matemática que é considerada “A Ciência que une a clareza do raciocínio à síntese da linguagem”, originou-se do convívio social, das trocas, da contagem, com caracter prático, utilitário e empírico;
- A Estatística é um ramo da Matemática que teve origem semelhante;
- Desde a Antigüidade vários povos já registravam o número de habitantes, de nascimento, de óbitos, faziam estimativas de riquezas individuais e sociais, etc;
- Na idade média colhiam-se informações, geralmente com a finalidade tributária;
- A partir do século XVI começaram a surgir às primeiras análises de fatos sociais, como batizados, casamentos, funerais, originando as primeiras tábuas e tabelas e os primeiros números relativos;
- No século XVII o estudo de tais fatos foi adquirindo proporções verdadeiramente científicas;
- Godofredo Achenwall, batizou a nova ciência (ou método) com o nome de ESTATÍSTICA, determinando assim o seu objetivo e suas relações com a ciência.

1.2. MÉTODO

Existem várias definições para métodos, Lakatos e Marconi (1982:39-40) mencionaram diversas definições, entre elas:

- Método é o “caminho pelo qual se chega a um determinado resultado...” (Hegemberg, 1976: II-115)
- Método é “um procedimento regular, explícito e passível de ser repetido para conseguirmos alguma coisa, seja material ou conceitual” (Bunge 1980: 19).

1.3. A ESTATÍSTICA

A definição de estatística não é única, a estatística abrange muito mais do que um simples traçado de gráficos e cálculos de medidas. Uma definição seria:

A estatística é uma coleção de métodos para planejar experimentos, obter dados e organizá-los, resumi-lo, analisá-los interpretá-los e deles extrair conclusões.

1.4. O MÉTODO ESTATÍSTICO

Dois métodos científicos podemos destacar: o método Experimental e o Método Estatístico.

O método experimental consiste em manter constante todas as causas (fatores) menos uma e variar esta causa de modo que o pesquisador possa descobrir seus efeitos caso existam.

O método estatístico diante da impossibilidade de se manter causas constantes, admite todas essas causas presentes variando-as registrando essa variação e procurando determinar no resultado final que influências cabem a cada uma delas.

RESUMO DA PROFISSÃO

O Estatístico promove o levantamento de pesquisas estatísticas em suas aplicações técnicas e científicas, investigando, elaborando e testando métodos matemáticos e sistema de amostragem, bem como coletando, analisando e interpretando os dados relacionados com os fenômenos estatísticos, e ainda estudando e renovando a metodologia estatística a fim de estabelecer a sua evolução e desenvolvimento.

ALGUMAS ESPECIALIZAÇÕES

Vinculam-se aos campos profissionais que exigem ou permitem o exercício do estatístico. Resultam da prática profissional e decorrem quase sempre da demanda decorrente no mercado de trabalho.

- Demografia
- Bioestatística
- Estatístico Matemático
- Estatístico de Estatística Aplicada, Etc.

CARGOS PROCURADOS

- Estatístico
- Estatístico Matemático
- Estatístico de Estatística Aplicada

1.5. A NATURZA DA ESTATÍSTICA

Podemos descrever duas variáveis para um estudo:

VARIÁVEIS QUALITATIVAS – (ou dados categóricos) podem ser separados em diferentes categorias, atributos, que se distinguem por alguma característica não numérica.

VARIÁVEIS QUANTITATIVAS – consistem em números que representam contagens ou medidias. Divide-se em:

VARIÁVEIS QUANT. DISCRETAS – resultam de um conjunto finito, enumerável de valores possíveis. Ex: número de filhos.

VARIÁVEIS QUANT. CONTÍNUAS – resultam de números infinitos de valores possíveis que podem ser associados a pontos em uma escala contínua. Ex: peso, altura.

Medida de Desobediência

Como coletar dados sobre algo que não se apresente mensurável, como o nível de desobediência do povo? O psicólogo Stanley Milgran planejou o seguinte experimento: Um pesquisador determinou que um voluntário acionasse um painel de controle que dava choques elétricos crescentemente dolorosos em uma terceira pessoa. Na realidade, não eram dados choques e a terceira pessoa era um ator. O voluntário começou com 15 volts e foi orientado a aumentar os choques de 15 em 15 volts. O nível de desobediência era o ponto em que a pessoa se recusava a aumentar a voltagem. Surpreendentemente, dois terços dos voluntários obedeceram às ordens mesmo que o ator gritasse e simulasse um ataque cardíaco.

Texto extraído do livro: Tiola, Mario F. **Introdução à Estatística.** 7ª ed. Rio de Janeiro – RJ. LTC. 1999.

1.6. USOS E ABUSOS DA ESTATÍSTICA

USOS DA ESTATÍSTICA

As Aplicações da estatística se desenvolveram de tal forma que, hoje, praticamente todo o campo de estudo se beneficia da utilização de métodos estatísticos. Os fabricantes fornecem melhores produtos a custos menores através de técnicas de controle de qualidade. Controlam-se doenças com o auxilio de análises que antecipam epidemias. Espécies ameaçadas são protegidas por regulamentos e leis que reagem a estimativas estatísticas de modificação de tamanho da população. Visando reduzir as taxas de casos fatais, os legisladores têm melhor justificativas para leis como as que regem a poluição atmosférica, inspeções de automóveis, utilização de cinto de segurança, etc.

ABUSOS DA ESTATÍSTICA

Não é de hoje que ocorrem abusos com a estatística. Assim é que , há cerca de um século, o estadista Benjamin Disraeli disse: "Há três tipos de mentiras: as mentiras, as mentiras sérias e as estatísticas". Já se disse também que "os números não mentem; mas os mentirosos forjam os números" (*Figures don't lie; liars figure*) e que "se torturarmos os dados por bastante tempo, eles acabam por admitir qualquer coisa". O historiador Andrew Lang disse que algumas pessoas usam a estatística "como um bêbado utiliza um poste de iluminação – para servir de apoio e não para iluminar". Todas essa afirmações se referem aos abusos da estatística quando os dados são apresentados de forma enganosa. Eis alguns exemplos das diversas maneiras como os dados podem ser distorcidos.

- Pequenas amostras
- Números imprecisos
- Estimativas por suposição
- Porcentagens distorcidas
- Cifras parciais
- Distorções deliberadas
- Perguntas tendenciosas
- Gráficos enganosos
- Pressão do pesquisador
- Más amostras

Os motoristas mais Idosos são mais Seguros do que os mais Moços?

A American Association of Retired People – AARP (Associação Americana de Aposentados) alega que os motoristas mais idosos se envolvem em menor número de acidentes do que os mais jovens. Nos últimos anos, os motoristas com 16-19 anos de idades causaram cerca de 1,5 milhões de acidentes em comparação com apenas 540.000 causados por motoristas com 70 anos ou mais, de forma que a alegação da AARP parece válida. Acontece, entretanto que os motoristas mais idosos não dirigem tanto quanto os mais jovens. Em lugar de considerar apenas o *número de acidentes*, devemos examinar também as taxas de acidentes. Eis as taxas de acidentes por 100 milhões de milhas percorridas: 8,6 para motoristas com idade de 16 a 19, 4,6 para os com idade de 75 a 79, 8,9 para os com idade 80 a 84 e 20,3 para os motoristas com 85 anos de idade ou mais. Embora os motoristas mais jovens tenham de fato o maior número de acidentes, os mais velhos apresentam as mais altas taxas de acidente.

Texto extraído do livro: Tiola, Mario F. **Introdução à Estatística.** 7ª ed. Rio de Janeiro – RJ. LTC. 1999.

1.7. ESTATÍSTICA DEDUTIVA E INDUTIVA

A estatística dedutiva também conhecida como Descritiva se encarrega de descrever o conjunto de dados desde a elaboração da pesquisa até o cálculo de determinada medida.

A estatística Indutiva ou inferencial está relacionada a incerteza. Inicia-se no cálculo das Probabilidades e se desenvolve por todo a área da inferência.

UNIDADE I – CONCEITOS INICIAIS EM ESTATÍSTICA

DEFINIÇÕES:

POPULAÇÃO: É um conjunto de indivíduos ou objetos que apresentam pelo menos uma característica em comum.

CENSO – é a coleção de dados relativos a todos os elementos da população.

AMOSTRA: Considerando a impossibilidade, na maioria das vezes do tratamento de todos os elementos da população, necessitaremos de uma parte representativa da mesma. A esta porção da população chamaremos de amostra.

ESTATÍSTICA: é a medida numérica que descreve uma característica da amostra.

PARÂMETRO – é a medida numérica que descreve uma característica da população.

RAMOS DA ESTATÍSTICA

A estatística possui três ramos principais:

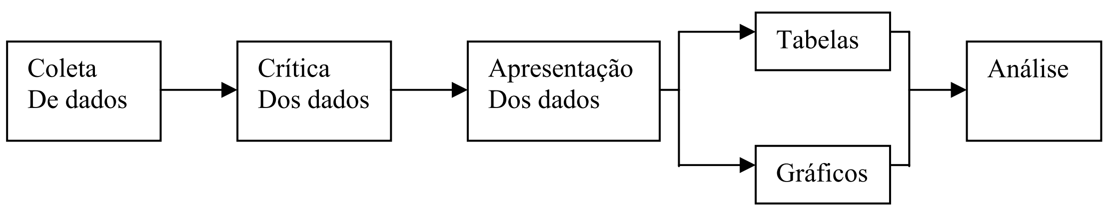
ESTATÍSTICA DESCRITIVA: envolve a organização e sumarização dos dados através de metodologias simples;

TEORIA DA PROBABILIDADE: que proporciona uma base racional para lidar com situações influenciadas por fatores que envolvem o **acaso.**

TEORIA DA INFERÊNCIA: que envolve a análise e interpretações da amostra.

ESTATÍSTICA DESCRITIVA

A Estatística Descritiva pode ser resumida no diagrama a seguir:



COLETA DOS DADOS:

Após a definição do problema a ser estudado e o estabelecimento do planejamento da pesquisa (forma pela qual os dados serão coletados; cronograma das atividades; custos envolvidos; exame das informações disponíveis; delineamento da amostra, etc.), o passo seguinte é a coleta dos dados, que consiste na busca ou compilação dos dados das variáveis, componentes do fenômeno a ser estudado.

A coleta dos dados é direta quando os dados são obtidos diretamente da fonte originária, como no caso da empresa que realiza uma pesquisa para saber a preferência dos consumidores pela sua marca.

A coleta dos dados é indireta quando é inferida a partir dos elementos conseguidos pela coleta direta.

CRÍTICA DOS DADOS

A revisão crítica dos dados procede com a finalidade de suprimir os valores estranhos ao levantamento, os quais são capazes de provocar futuros enganos.

APRESENTAÇÃO DOS DADOS

Convém organizarmos o conjunto de dados de maneira prática e racional. Tal organização denomina-se Série Estatística (que será abordado na próxima unidade). Sua apresentação pode ocorrer por meio de Tabelas e/ou Gráficos.

TÉCNICAS DE AMOSTRAGEM

As regras de Amostragem podem ser classificadas em duas categorias gerais:

PROBABILÍSTICA - São amostragem em que a seleção é **aleatória** de tal forma que cada elemento tem igual probabilidade de ser sorteado para a amostra.

NÃO-PROBABILISTICAS OU INTENCIONADAS - São amostragem em que há uma escolha deliberada dos elementos da amostra.

TIPOS DE AMOSTRAGEM

AMOSTRAGEM ALEATÓRIA SIMPLES

Também conhecida por amostragem ocasional, acidental, casual, randômica, etc. A amostragem simples ao acaso destaca-se por ser um processo de seleção bastante fácil e muito usado. Neste processo, todos os elementos da população têm **igual probabilidade** de serem escolhidos, desde o início até completo processo de coleta.

PROCEDIMENTO

1. Devemos enumerar todos os elementos da população
2. Devemos efetuar sucessivos sorteios com reposição até completar o tamanho da amostra (n)

Para realizarmos este sorteio devemos fazer uso das "tábuas de números aleatórios" (veja página seguinte). Estas apresentam os dígitos de 0 a 9 distribuídos aleatoriamente.

EXEMPLO:

Supor que nós tenhamos uma população com 1.000 elementos, que numeramos de 000 a 999, para selecionarmos uma amostra aleatória, de 200 elementos, basta escolhermos uma posição de qualquer linha e extrairmos conjuntos de três algarismos, até completarmos os 200 elementos da amostra. O processo termina quando for sorteado o elemento 200. Se o número sorteado não existia na população simplesmente não o consideramos, e prosseguimos com o processo.

AMOSTRAGEM SISTEMÁTICA

Trata-se de uma variação da amostragem simples ao acaso, muito conveniente quando a população está **naturalmente ordenada**, como fichas em um fichário, listas telefônicas etc. Requer uma lista dos itens da população, e, assim, padece das mesmas restrições já mencionadas na aleatória ao acaso. Se os itens da lista não se apresentarem numa ordem determinada à amostragem Sistemática pode dar uma amostra realmente aleatória.

PROCEDIMENTO

Sejam os seguintes elementos:

- *N:* tamanho da população;
- *n:* tamanho da amostra.

Então, calcula-se o intervalo de amostragem através da razão $a = \frac{N}{n}$ (onde a é o inteiro mais próximo). Sorteia-se, utilizando a tábua de números aleatórios, um número x entre **1** e ***a*** formando-se a amostra dos elementos correspondentes ao conjunto de números:

$$x;\ x+a; x+2a; \ldots;\ x+(n-1)a.$$

EXEMPLO: Seja $N = 500$, $n = 50$. Então $a = \frac{500}{50} = 10$

Sorteia-se um número de 1 a 10. Seja 3 (x = 3) o número sorteado. Logo, os elementos numerados por 3;13;23;33;... serão os componentes da amostra.

AMOSTRAGEM ESTRATIFICADA

No caso de possuir uma população com uma certa característica heterogênea, na qual podemos distinguir subpopulações mais ou menos homogêneas, denominadas de estratos, podemos usar a amostragem estratificada.

Estratificar uma população em L subpopulações denominada estratos, tais que:

$$n_1 + n_2 + \ldots + n_L = n$$

Onde os estratos são **mutuamente exclusivos**.

Após a determinação dos estratos, seleciona-se uma amostra aleatória de cada sub-população.

Se as diversas sub-amostras tiverem tamanhos proporcionais ao respectivo número de elementos nos estratos, teremos a estratificação proporcional.

Stevenson, William J. *Estatística aplicada à administração.* Harper & Row do Brasil, São Paulo, 1986, p.165

Tabela 6.1 Números Aleatórios

3690	2492	7171	7720	6509	7549	2330	5733	4730
0813	6790	6858	1489	2669	3743	1901	4971	8280
6477	5289	4092	4223	6454	7632	7577	2816	9202
0772	2160	7236	0812	4195	5589	0830	8261	9232
5692	9870	3583	8997	1533	6466	8830	7271	3809
2080	3828	7880	0586	8482	7811	6807	3309	2729
1039	3382	7600	1077	4455	8806	1822	1669	7501
7227	0104	4141	1521	9104	5563	1392	8238	4882
8506	6348	4612	8252	1062	1757	0964	2983	2244
5086	0303	7423	3298	3979	2831	2257	1508	7642
0092	1629	0377	3590	2209	4839	6332	1490	3092
0935	5565	2315	8030	7651	5189	0075	9353	1921
2605	3973	8204	4143	2677	0034	8601	3340	8383
7277	9889	0390	5579	4620	5650	0210	2082	4664
5484	3900	3485	0741	9069	5920	4326	7704	6525
6905	7127	5933	1137	7583	6450	5658	7678	3444
8387	5323	3753	1859	6043	0294	5110	6340	9137
4094	4957	0163	9717	4118	4276	9465	8820	4127
4951	3781	5101	1815	7068	6379	7252	1086	8919
9047	0199	5068	7447	1664	9278	1708	3625	2864
7274	9512	0074	6677	8676	0222	3335	1976	1645
9192	4011	0255	5458	6942	8043	6201	1587	0972
0554	1690	6333	1931	9433	2661	8690	2313	6999
9231	5627	1815	7171	8036	1832	2031	6298	6073
3995	9677	7765	3194	3222	4191	2734	4469	8617
2402	6250	9362	7373	4757	1716	1942	0417	5921
5295	7385	5474	2123	7035	9983	5192	1840	6176
5177	1191	2106	3351	5057	0967	4538	1246	3374
7315	3365	7203	1231	0546	6612	1038	1425	2709
5775	7517	8974	3961	2183	5295	3096	8536	9442
5500	2276	6307	2346	1285	7000	5306	0414	3383
3251	8902	8843	2112	8567	8131	8116	5270	5994
4675	1435	2192	0874	2897	0262	5092	5541	4014
3543	6130	4247	4859	2660	7852	9096	0578	0097
3521	8772	6612	0721	3899	2999	1263	7017	8057
5573	9396	3464	1702	9204	3389	5678	2589	0288
7478	7569	7551	3380	2152	5411	2647	7242	2800
3339	2854	9691	9562	3252	9848	6030	8472	2266
5505	8474	3167	8552	5409	1556	4247	4652	2953
6381	2086	5457	7703	2758	2963	8167	6712	9820

Fonte: Donald B. Owen, *Handbook of Statistical Tables*, Reading, Mass: Addison-Wesley, 1962. (Cortesia de U.S. Energy Research and Development Adm.)

EXERCÍCIOS

1. População ou universo é:
   a) Um conjunto de pessoas;
   b) Um conjunto de elementos quaisquer
   c) Um conjunto de pessoas com uma característica comum;
   d) Um conjunto de elementos com pelo menos uma característica em comum;
   e) Um conjunto de indivíduo de um mesmo município, estado ou país.

2. Uma parte da população retirada para analisá-la denomina-se:
   a) Universo;
   b) Parte;
   c) Pedaço;
   d) Dados Brutos;
   e) Amostra.

3. A parte da estatística que se preocupa somente com a descrição de determinadas características de um grupo, sem tirar conclusões sobre um grupo maior denomina-se:
   a) Estatística de População;
   b) Estatística de Amostra;
   c) Estatística Inferencial
   d) Estatística Descritiva;
   e) Estatística Grupal.

4. Diga qual tipo de variáveis estamos trabalhando nos casos abaixo:
   a. No. de inscrições no Seguro Social
   b. No. de passageiros no ônibus da linha Rio-São Paulo
   c. Escolaridade
   d. Peso Médio dos Recém Nascidos
   e. Altitude acima do nível do mar
   f. Uma pesquisa efetuada com 1015 pessoas indica que 40 delas são assinantes de um serviço de computador on-line
   g. Cada cigarro Camel tem 16,13mg de alcatrão
   h. O radar indique que Nolan Ryan rebateu a ultima bola a 82,3mi/h
   i. O tempo gasta para uma pessoa fazer uma viagem de carro de Brasília até Belo Horizonte é de aproximadamente 8:00h a uma velocidade média de 93,75km/hs

5. Classifique as seguintes variáveis:
   a) Cor dos olhos
      i) Qualitativa;
      ii) Qualitativa discreta;
      iii) Quantitativa contínua;
      iv) Quantitativa discreta;
      v) Qualitativa contínua.

b) Número de filhos de um casal:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

c) Peso de um indivíduo:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

d) Altura de um indivíduo:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

e) Número de alunos de uma escola:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

f) Tipo sangüíneo:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

g) Fator RH:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

h) Valor obtido na face superior de um dado:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

i) Sexo:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

j) Resultado da extração da loteria Federal:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

k) Comprimento de um seguimento de reta:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

l) Área de um Círculo:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

m) Raça:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

n) Quantidade de livro de uma biblioteca:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

o) Religião:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

p) Salário dos Empregados de uma empresa:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

q) Estado Civil:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

r) Profissão:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

s) Volume de água contido numa piscina:
   i) Qualitativa;
   ii) Qualitativa discreta;
   iii) Quantitativa contínua;
   iv) Quantitativa discreta;
   v) Qualitativa contínua.

6. Suponha que existem N = 1.000 fichas de pacientes das quais uma amostra aleatória de n = 20 deve ser selecionada. Determine que fichas devem ser escolhidas na amostra de tamanho n = 20. Diga que tipo de amostragem foi feito e como foram selecionadas as fichas.

7. Suponha que uma pesquisa de opinião pública deve ser realizada em um estado que tem duas grandes cidades e uma zona rural. Os elementos na população de interesse são todos os homens e mulheres do estado com idade acima de 21 anos. Diga que tipo de amostragem utilizará?

8. Serviço florestal do estado está conduzindo um estudo das pessoas que usam as estruturas de um camping operado por ele. O estado tem duas áreas de camping, uma localizada nas montanhas e outra localizada ao longo da costa. O serviço florestal deseja estimar o número médio de pessoas por acampamento e a proporção de acampamento ocupada por pessoas de fora do estado, durante o fim de semana em particular, quando se espera que todos os acampamentos estejam ocupados. Sugira um plano amostral e explique rapidamente como devem ser feitos.

9. Um médico está interessado em obter informação sobre o número médio de vezes em que 15.000 especialistas prescreveram certa droga no ano anterior (N = 15.000). Deseja-se obter n = 1.600. Que tipo de amostragem você sugeriria e por que?

10. Um hematologista deseja fazer uma nova verificação de uma amostra de n = 10 dos 854 espécimes de sangue analisados por um laboratório médico em um determinado mês. Que tipo de amostragem você sugeriria e por que?

11. Um repórter da revista Business Week obtém uma relação numerada de 1.000 empresas com maiores de cotações de ações na bolsa. Ele entrevistará 100 gerentes gerais das empresas correspondentes a esta amostra. Que tipo de amostragem você sugeriria e por que?

12. Comente rapidamente sobre a pesquisa abaixo

> "Um relatório patrocinado pela Flórida Citrus Comission concluiu que os níveis de colesterol podem ser reduzidos mediante ingestão de produtos cítricos".

Por que razão a conclusão poderia ser suspeita

13. Dada uma população com seis elementos, A, B, C, D, E e F, explique como você faria para obter, dessa população, uma amostra aleatória simples com três elementos.

14. Descreva uma forma de se obter uma amostra sistemática com 10 elementos de uma população com tamanho 100.

15. Explique a forma de se obter uma amostragem estratificada dos empregados de uma firma, considerando que existem empregados de escritório, de oficina e representantes da mesma.

16. Imagine que se pretenda fazer um levantamento de opinião pública para verificar se as pessoas são contra ou a favor do uso gratuito de ônibus pelos idosos. Pense em três maneiras distintas de elaborar uma pergunta que induza a resposta positiva, outra que induza a resposta negativa e uma outra que não ocorra nenhum tipo de tendência na resposta.

17. Identifique o tipo de amostragem utilizado para cada uma das situações abaixo:

    a. Quando escreveu *Woman in Love: A Cultural Revolution*, a autora Shere Hite baseou suas conclusões em 4.500 respostas a 100.000 questionários distribuídos a mulheres.

    b. Uma psicóloga da Universidade de Nova York faz uma pesquisa sobre alguns alunos selecionados aleatoriamente de todas as 20 turmas que participaram desta pesquisa.

    c. Um sociólogo da Universidade Charleston seleciona 12 homens e 12 mulheres de cada uma de quatro turmas de inglês.

    d. A empresa Sony seleciona cada 200º CD de sua linha de produção e faz um teste de qualidade rigoroso.

    e. Um cabo eleitoral escreve o nome de cada senador dos EUA em cartões separados, mistura-os e extrai 10 nomes.

    f. Gerente comercial da America OnLine testa uma nova estratégia de vendas selecionando aleatoriamente 250 consumidores com renda inferior a US$50.000,00 e 250 consumidores com renda de ao menos de US$50.000,00.

    g. O programa Planned Parenthood (Planejamento Familiar) pesquisa 500 homens e 500 mulheres sobre seus pontos de vista sobre o uso de anticoncepcionais.

    h. Um repórter da revista Business Week Entrevista todo o 50º gerente geral constante da relação das 1.000 empresas com maior cotação de suas ações.

    i. Um repórter da revista Business Week obtém uma relação numerada das 1.000 empresas com maior cotação de ações na bolsa, utiliza um computador para gerar 20 números aleatórios e então entrevista gerentes gerais das empresas correspondentes aos números extraídos.

UNIDADE II - NORMAS PARA CONSTRUÇÃO DE TABELAS

TABELAS ESTATÍSTICAS

Um dos objetivos da estatística é sintetizar os valores que uma ou mais variáveis podem assumir, para que tenhamos uma visão global da variação das mesmas.

Tabela é uma maneira de apresentar de forma resumida um conjunto de dados

ELEMENTOS DE UMA TABELA

A tabela se apresenta da seguinte forma:

TÍTULO DA TABELA

CORPO DA TABELA

RODAPÉ

EXEMPLO:

Tabela 1 – Produção de Café Brasil – 1991 a 1995

Anos	Produção (1.000 t)
1991	2.535
1992	2.666
1993	2.122
1994	3.750
1995	2.007

Fonte: IBGE

TÍTULO DA TABELA:

Conjunto de informações, as mais completas possíveis, respondendo às perguntas: **O que?, Quando?** e **Onde?**, Localizado no topo da tabela, além de conter a palavra "TABELA" e sua respectiva numeração.

CORPO DA TABELA:

É o conjunto de Linhas e Colunas que contém informações sobre a variável em estudo.

a) **Cabeçalho da Coluna** – Parte superior da tabela que especifica o conteúdo das colunas;

b) **Coluna Indicadora** – Parte da tabela que especifica o conteúdo das linhas;

c) **Linhas** – retas imaginárias que facilitam a leitura, no sentido horizontal, de dados que se inscrevem nos seus cruzamentos com as linhas;

d) **Casa** ou **Célula** – espaço destinado a um só número;

e) **Total** – deve ser SEMPRE destacado de alguma forma;

f) **Laterais da tabela** – não devem ser fechadas. Caso as feche, passa a ser chamada de "QUADRO".

g) **Número** – preferencialmente utilizar separador de 1000 (por exemplo: 1.854.985 ao invés de 1854985).

Há ainda a considerar os elementos complementares da tabela, que são a **fonte**, as **notas**, e as **chamadas**, localizadas, de preferência, no rodapé.

a) **Fonte** – identifica o responsável (pessoa física ou jurídica) ou responsável pelos dados numéricos;

b) **Notas** – é o texto que irá esclarecer o conteúdo estudado, que poderá ser de caráter geral ou específico de uma tabela;

c) **Chamadas** – símbolo remissivo atribuído a algum elemento de uma tabela que necessita de uma nota específica.

SINAL CONVENCIONAL:

A substituição de uma informação da tabela poderá ser feita pelos sinais abaixo:

a) - dado numérico igual a zero;

b) ... Quando não temos os dados;

c) ? Quando temos dúvida na informação;

d) 0 quando o valor for muito pequeno.

SÉRIES ESTATÍSTICAS

Introdução

Uma vez que os dados foram coletados, muitas vezes o conjunto de valores é extenso e desorganizado, e seu exame requer atenção, pois há o risco de se perder a visão global do fenômeno analisado. Para que isto não ocorra faz-se necessário reunir os valores em tabelas convenientes, facilitando sua compreensão.

Além da apresentação do conjunto de valores na forma tabulada, tem-se também a forma gráfica, que por sua vez, representa uma forma mais útil e elegante de representar o conjunto dos valores.

Qualquer que seja a forma de representação do conjunto de valores, desde de que não haja alterações em seus valores iniciais, quer seja o de caracterização de um conjunto, ou de comparação com outros semelhantes ou ainda o de previsão de valores possíveis, facilitará sua compreensão de qualquer estudo. É o caso da série estatística.

Definição de Série Estatística

Uma série estatística define-se como toda e qualquer coleção de dados estatísticos referidos a uma mesma ordem de classificação: QUANTITATIVA. Em um sentido mais amplo, SÉRIE é uma seqüência de números que se refere a uma certa variável.

Caso estes números expressem dados estatísticos a série é chamada de **série estatística**. Em um sentido mais restrito, diz-se que uma **série estatística** é uma sucessão de dados estatísticos referidos a caracteres quantitativos.

Para diferenciar uma série estatística de outra, temos que levar em consideração três fatores:

◊ A ÉPOCA (fator temporal ou cronológico) a que se refere o fenômeno analisado;

◊ O LOCAL (fator espacial ou geográfico) onde o fenômeno acontece;

◊ O FENÔMENO (espécie do fator ou fator específico) que é descrito.

Tipos de Séries Estatísticas

São quatro os tipos de séries estatísticas conforme a variação de um dos fatores:

- **SÉRIE TEMPORAL**

A série temporal, igualmente chamada série cronológica, histórica, evolutiva ou marcha, identifica-se pelo caráter variável do fator cronológico. Assim deve-se ter:

VARIÁVEL: a época

FIXO: o local e o fenômeno

- **SÉRIE GEOGRÁFICA**

Também denominadas séries territoriais, espaciais ou de localização, esta série apresenta como elemento ou caráter variável somente o fator local. Assim:

VARIÁVEL: o local

FIXO: a época e o fenômeno

- **SÉRIE ESPECÍFICA**

A série específica recebe também outras denominações tais como série categórica ou série por categoria. Agora o caráter variável é o fenômeno.

VARIÁVEL: o fenômeno

FIXO: a época e o local

- **DISTRIBUIÇÃO DE FREQÜÊNCIA**

Neste caso **todos** os elementos (época, local e fenômeno) são fixos. Embora fixo, o fenômeno apresenta-se agora através de graduações, isto é, os dados referentes ao fenômeno que se está representando são reunidos de acordo com a sua magnitude. Normalmente os problemas de tabulação são enquadrados neste tipo de série, que iremos estudar com maior detalhe mais adiante neste curso.

<u>Proporção, Porcentagem e Razão.</u>

Introdução

Do ponto de vista estatístico, estas podem ser consideradas como medidas muito simples que permitem estabelecer comparações entre diversos grupos.

- **Proporção**

Considere um número de empregados que foi distribuído em quatro repartições de uma certa empresa de acordo com sua função. Estas repartições são mutuamente exclusivas (cada pessoa somente poderá ser alocada em uma única repartição) e exaustivas (todas as pessoas deverão ser alocadas).

Em termos simbólicos podemos escrever:

N_1 = número de pessoas alocadas na repartição 1

N_2 = número de pessoas alocadas na repartição 2

N_3 = número de pessoas alocadas na repartição 3

N_4 = número de pessoas alocadas na repartição 4

$N = N_1 + N_2 + N_3 + N_4$ = número total de empregados

Neste caso, a proporção de empregados pertencentes à primeira repartição é determinada mediante o cálculo do quociente $\frac{N_1}{N}$; para as demais repartições segue o mesmo procedimento: $\frac{N_2}{N}$, $\frac{N_3}{N}$ e $\frac{N_4}{N}$.

Note que o valor de uma proporção não pode exceder a unidade, e que a soma de todas as proporções será **sempre** igual à unidade. Assim,

$$\frac{N_1}{N} + \frac{N_2}{N} + \frac{N_3}{N} + \frac{N_4}{N} = \frac{N}{N} = 1$$

Exemplo:

Tabela 01. Número de empregados contratados (consultores) e com carteira assinada em dois órgãos públicos

EMPREGADO	ÓRGÃO PÚBLICO 1	ÓRGÃO PÚBLICO 2
CONSULTOR:		
TEMPO INTEGRAL	580	680
MEIO EXPEDIENTE	430	1.369
CARTEIRA ASSINADA	4.810	10.811
TOTAL	5.820	12.860

FONTE: Departamento de Recursos Humanos destes Órgãos Públicos

Não é simples raciocinar em termos absolutos e dizer qual dos dois órgãos públicos conta com maior número de empregados consultores em suas duas modalidades de expedientes porque o número total de empregados difere muito entre si. Por outro lado, a comparação direta pode ser estabelecida rapidamente, se os dados forem expressos em proporções.

A proporção de consultores com tempo integral no órgão público 1 é:

$$\frac{N_1}{N} = \frac{580}{5.820} = 0{,}099 \cong 0{,}1$$

E no órgão público 2, seguindo o mesmo raciocínio temos:

$$\frac{N_1}{N} = \frac{680}{12.860} = 0{,}0528 \cong 0{,}053$$

Note que, em números absolutos, estes valores são muito próximos (580 e 680). Entretanto, o órgão público 2 apresenta uma proporção inferior de consultores com tempo integral.

Analogamente, fazendo os cálculos para ambos os órgãos públicos, têm:

- ÓRGÃO PÚBLICO 1
  - Consultores com ½ expediente: $\frac{N_2}{N} = \frac{430}{5.820} = 0{,}0738 \cong 0{,}074$
  - Carteira assinada: $\frac{N_3}{N} = \frac{4.810}{5.820} = 0{,}8264 \cong 0{,}826$
- ÓRGÃO PÚBLICO 2
  - Consultores com ½ expediente: $\frac{N_2}{N} = \frac{1.369}{12.860} = 0{,}1064 \cong 0{,}106$
  - Carteira assinada: $\frac{N_3}{N} = \frac{10.811}{12860} = 0{,}8406 \cong 0{,}841$

Assim, temos a seguinte tabela de proporções:

Tabela 02. Proporção de empregados contratados (consultores) e com carteira assinada em dois órgãos públicos

EMPREGADO	ÓRGÃO PÚBLICO 1	ÓRGÃO PÚBLICO 2
CONSULTOR:		
TEMPO INTEGRAL	0,100	0,053
MEIO EXPEDIENTE	0,074	0,106
CARTEIRA ASSINADA	0,826	0,841
TOTAL	1	1

FONTE: Departamento de Recursos Humanos destes Órgãos Públicos

- **Porcentagem**

As porcentagens são obtidas a partir do cálculo das proporções, simplesmente multiplicando-se o quociente obtido por 100. A palavra *porcentagem* significa, portanto, "por cem". Uma vez que a soma das proporções é igual a 1, a soma das porcentagens é igual a 100, a menos que as categorias não sejam mutuamente exclusivas e exaustivas.

Exemplo: Utilizando os dados do exemplo anterior e multiplicando as proporções por 100 teremos a seguinte tabela:

Tabela 03. Percentual de empregados contratados (consultores) e com carteira assinada em dois órgãos públicos

EMPREGADO	ÓRGÃO PÚBLICO 1		ÓRGÃO PÚBLICO 2	
	ABSOLUTO	RELATIVO (%)	ABSOLUTO	RELATIVO (%)
CONSULTOR:				
TEMPO INTEGRAL	580	10,0	680	5,3
MEIO EXPEDIENTE	430	7,4	1.369	10,6
CARTEIRA ASSINADA	4.810	82,6	10.811	84,1
TOTAL	5.820	100	12.860	100

FONTE: Departamento de Recursos Humanos destes Órgãos Públicos

As porcentagens e proporções, em Estatística, têm como principal finalidade estabelecer comparações relativas. Como um outro exemplo, as vendas de duas empresas foram as seguintes em dois anos consecutivos:

Tabela 4. Faturamento anual das Empresas A e B em 1994 e 1995 dados em números absoluto e relativo (%)

EMPRESA	FATURAMENTO (por 1.000 reais)		CRESCIMENTO ABSOLUTO	CRESCIMENTO RELATIVO (%)
	1994	1995		
A	2.000	3.000	1.000	50
B	20.000	25.000	5.000	25

FONTE: Departamento de Finanças das Empresas A e B

Em valores absolutos, a empresa **B** teve um crescimento no faturamento maior que a empresa **A**. Contudo, na realidade, comparando estes valores em termos percentuais, a empresa **A** foi a que apresentou um desempenho superior (crescimento de 50% na empresa **A** e de 25% na empresa **B**).

- **Razão**

A razão de um número **A** em relação a outro número **B** define-se como "**A** dividido por **B**" A quantidade precedente é posta no numerador e a seguinte, no denominador.

Exemplo: Através de uma pesquisa realizada em uma certa cidade, descobriu-se que, das pessoas entrevistadas, 300 se manifestaram a favor a uma determinada medida adotada pela prefeitura local, 400 contra e 70 eram indiferentes. Neste caso, a razão daquelas pessoas contra a medida para aquelas a favor foi de:

$$\frac{400}{300} \text{ ou } \frac{4}{3} \text{ ou } 4{:}3 \text{ ou } 1{,}33 \text{ para } 1$$

E a razão daquelas a favor e contra para aquelas indiferentes foi de:

$$\frac{(400+300)}{70} \text{ ou } \frac{70}{7} \text{ ou } 70{:}7 \text{ ou } 10 \text{ para } 1$$

EXERCÍCIOS

1. Uma série estatística é denominada <u>evolutiva</u> quando?
   a) O elemento variável é o tempo;
   b) O elemento variável é o local;
   c) O elemento variável é a espécie;
   d) É o resultado da combinação de séries estatísticas de tipos diferentes;
   e) Os dados são agrupados em subintervalos do intervalo observado.

2. Uma série estatística é denominada <u>espacial</u> quando?
   f) O elemento variável é o tempo;
   g) O elemento variável é o local;
   h) O elemento variável é a espécie;
   i) É o resultado da combinação de séries estatísticas de tipos diferentes;
   j) Os dados são agrupados em subintervalos do intervalo observado.

3. Uma série estatística é denominada <u>cronológica</u> quando?
   a) O elemento variável é o tempo;
   b) O elemento variável é o local;
   c) O elemento variável é a espécie;
   d) É o resultado da combinação de séries estatísticas de tipos diferentes;
   e) Os dados são agrupados em subintervalos do intervalo observado.

4. Uma série estatística é denominada <u>categórica</u> quando?
   a) O elemento variável é o tempo;
   b) O elemento variável é o local;
   c) O elemento variável é a espécie;
   d) É o resultado da combinação de séries estatísticas de tipos diferentes;
   e) Os dados são agrupados em subintervalos do intervalo observado.

5. Uma série estatística é denominada <u>marcha</u> quando?
   a) O elemento variável é o tempo;
   b) O elemento variável é o local;
   c) O elemento variável é a espécie;
   d) É o resultado da combinação de séries estatísticas de tipos diferentes;
   e) Os dados são agrupados em subintervalos do intervalo observado.

6. Uma série estatística é denominada <u>geográfica</u> quando?
   a) O elemento variável é o tempo;
   b) O elemento variável é o local;
   c) O elemento variável é a espécie;
   d) É o resultado da combinação de séries estatísticas de tipos diferentes;
   e) Os dados são agrupados em subintervalos do intervalo observado.

7. Uma série estatística é denominada <u>composta</u> quando?
   a) O elemento variável é o tempo;

b) O elemento variável é o local;
c) O elemento variável é a espécie;
d) É o resultado da combinação de séries estatísticas de tipos diferentes;
e) Os dados são agrupados em subintervalos do intervalo observado.

8. Uma série estatística é denominada qualitativa quando?
   a) O elemento variável é o tempo;
   b) O elemento variável é o local;
   c) O elemento variável é a espécie;
   d) É o resultado da combinação de séries estatísticas de tipos diferentes;
   e) Os dados são agrupados em subintervalos do intervalo observado.

9. Uma série estatística é denominada específica quando?
   a) O elemento variável é o tempo;
   b) O elemento variável é o local;
   c) O elemento variável é a espécie;
   d) É o resultado da combinação de séries estatísticas de tipos diferentes;
   e) Os dados são agrupados em subintervalos do intervalo observado.

10. Uma série estatística é denominada mista quando?
    a) O elemento variável é o tempo;
    b) O elemento variável é o local;
    c) O elemento variável é a espécie;
    d) É o resultado da combinação de séries estatísticas de tipos diferentes;
    e) Os dados são agrupados em subintervalos do intervalo observado.

11. Uma série estatística é denominada Temporal quando?
    a) O elemento variável é o tempo;
    b) O elemento variável é o local;
    c) O elemento variável é a espécie;
    d) É o resultado da combinação de séries estatísticas de tipos diferentes;
    e) Os dados são agrupados em subintervalos do intervalo observado.

12. A representação tabular de dados no Brasil obedece às normas
    a) Da SUNAB;
    b) Da Receita Federal;
    c) Do IBGE;
    d) Do Governo Federal;
    e) Da Secretaria Municipal de Estatística.

13. De acordo com as normas para representação tabular de dados, quando o valor de um dado é zero, deve-se colocar na célula correspondente:
    a) Zero (0);
    b) Três pontos (...);
    c) Um traço horizontal (-)
    d) Um ponto de interrogação (?);
    e) Um ponto de exclamação (!).

14. De acordo com as normas para representação tabular de dados, quando o valor de um dado é não está disponível, deve-se colocar na célula correspondente.
    a) Zero (0);
    b) Três pontos (...);
    c) Um traço horizontal (-)
    d) Um ponto de interrogação (?);
    e) Um ponto de exclamação (!).

15. De acordo com as normas para representação tabular de dados, quando o valor de um dado é muito pequeno, para ser expresso com o número de casa decimais utilizadas ou com a unidade de medida utilizada, deve-se colocar na célula correspondente.
   a) Zero (0);
   b) Três pontos (...);
   c) Um traço horizontal (-)
   d) Um ponto de interrogação (?);
   e) Um ponto de exclamação (!).

16. De acordo com as normas para representação tabular de dados, quando há dúvida, na exatidão do valor de um dado, deve-se colocar na célula correspondente.
   a) Zero (0);
   b) Três pontos (...);
   c) Um traço horizontal (-)
   d) Um ponto de interrogação (?);
   e) Um ponto de exclamação (!).

17. Assinale a alternativa verdadeira
   a) Tanto a nota quanto a chamada são usadas para esclarecimento geral sobre um quadro e uma tabela.
   b) Tanto a nota quanto a chamada são usadas para esclarecer detalhes em relação à casa, linhas ou colunas de um quadro ou uma tabela.
   c) A nota é usada para esclarecer detalhes em relação a casas, linhas ou colunas enquanto a chamada é usada para um esclarecimento geral sobre um quadro ou uma tabela.
   d) A nota é usada para esclarecimento geral sobre um quadro ou tabela enquanto a chamada é usada para esclarecer detalhes em relação a casas, linhas ou colunas.
   e) Todas as afirmativas anteriores são falsas.

18. Para cada tabela abaixo, calcule a proporção e a porcentagem e responda às perguntas:

Tabela 01. Quociente de Inteligência (QI) de uma certa faculdade brasileira

QI	No. DE ALUNOS	PROPORÇÃO	PORCENTAGEM
092 |- 107	31		
107 |- 122	39		
122 |- 137	21		
137 |- 152	12		
152 |- 167	4		
TOTAL	107		

   a) Qual o nível de QI que possui a maior proporção/percentual? E a menor?
   b) Calcule e interprete as seguintes razões:
      i) Alunos com QI entre 92 e 122 (exclusive) para aqueles com QI entre 137 e 152 (exclusive).
      ii) Alunos com QI entre 107 e 152 (exclusive) para os demais.
      iii) Alunos com QI entre 92 e 107 (exclusive) para aqueles com QI entre 152 e 167 (exclusive).
      iv) Alunos com QI inferior a 122 para aqueles com QI maior ou igual a 137.

Tabela 02. Notas de candidatos de um certo concurso público realizado em uma cidade

NOTAS	FREQUÊNCIA	PROPORÇÃO	PORCENTAGEM
00|-20	20		
20|-40	65		
40|-60	230		
60|-80	160		
80|-100	25		
TOTAL	500		

a) Dado que a nota de corte seja de 60 pontos, qual a proporção/percentual dos candidatos que foram aprovados?

b) Calcule e interprete as seguintes razões:

   i) Candidatos com nota menor que 20 para aqueles com nota de 40 a 60 (exclusive).

   ii) Candidatos com nota menor que 40 para aqueles com nota mínima de 60.

   iii) Candidatos com nota de 40 a 60 (exclusive) para aqueles com nota igual ou superior a 80.

   iv) Candidatos com nota máxima de 40 para aqueles com nota maior ou igual a 60.

   v) Candidatos com nota de 20 a 60 (exclusive) para os demais.

Tabela 03. Área das Regiões Brasileiras

REGIÃO	ÁREA	PROPORÇÃO	PORCENTAGEM
NORTE	3.581.180		
NORDESTE	1.546.672		
SUDESTE	924.935		
SUL	577.723		
C.OESTE	1.879.455		
TOTAL	8.509.965		

a) Qual a região que ocupa a maior área do Brasil e qual é a sua proporção/porcentagem?

b) Calcule e interprete as seguintes razões:

   i) Área da região Norte para a da região Nordeste.

   ii) Área das regiões Norte e Nordeste para o da região Centro-Oeste.

   iii) Área da região Sudeste para o das regiões Sul e Centro-Oeste.

   iv) Área da região Norte para as demais.

UNIDADE III - NORMAS PARA CONSTRUÇÃO DE GRÁFICOS

- **Introdução**

Tem como finalidade:

- Representar os resultados de forma simples, clara e verdadeira.
- Demonstrar a evolução do fenômeno em estudo
- Observar a relação dos valores da série

- **Normas para construção de gráficos**

A disposição dos elementos é idêntica à das tabelas:

CABEÇALHO DO GRÁFICO
CORPO DO GRÁFICO
RODAPÉ

➢ TIPOS DE GRÁFICOS

GRÁFICO EM COLUNAS

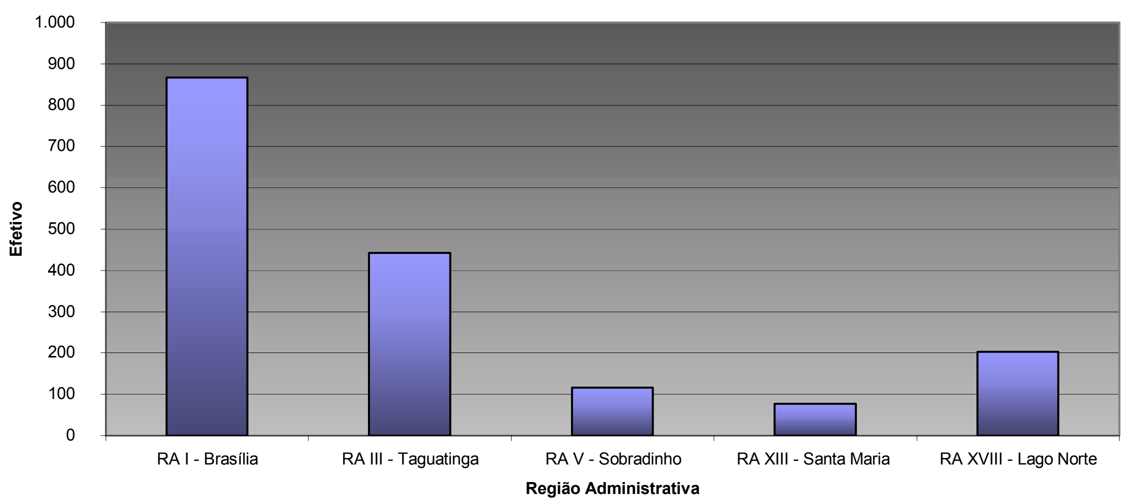
❖ Conjunto de retângulos dispostos verticalmente separados por um espaço.

Tabela 01. Efetivo do CBMDF em Cinco Regiões Administrativas do DF - 1998

Região Administrativa	Efetivo
RA I - Brasília	867
RA III - Taguatinga	443
RA V - Sobradinho	116
RA XIII - Santa Maria	77
RA XVIII - Lago Norte	203
Total	1.706

FONTE: Banco de Dados do Distrito Federal – 1998
NOTAS: Os efetivos especializados (emergência médica, incêndio florestal e guarda e segurança) estão alocados nas regiões administrativas.

Gráfico 01. Efetivo do CBMDF em algumas Regiões Administrativas do DF - 1998



Fonte: Tabela 01

GRÁFICOS EM BARRAS

❖ Semelhante ao gráfico em colunas, porém os retângulos são dispostos horizontalmente.

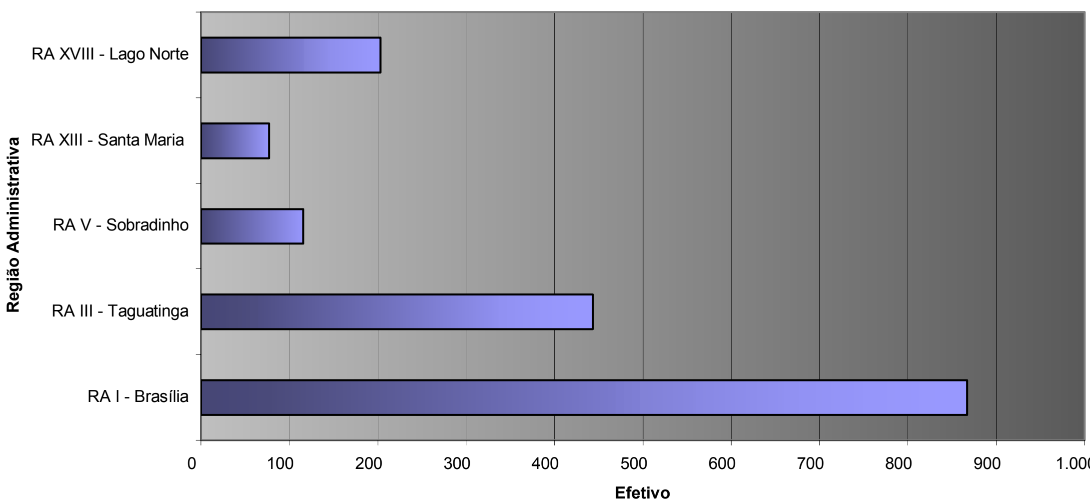
Tabela 02. Efetivo do CBMDF em Cinco Regiões Administrativas do DF - 1998

Região Administrativa	Efetivo
RA I - Brasília	867
RA III - Taguatinga	443
RA V - Sobradinho	116
RA XIII - Santa Maria	77
RA XVIII - Lago Norte	203
Total	1.706

FONTE: Banco de Dados do Distrito Federal – 1998

NOTAS: Os efetivos especializados (emergência médica, incêndio florestal e guarda e segurança) estão alocados nas regiões administrativas.

Gráfico 02. Efetivo do CBMDF em algumas Regiões Administrativas do DF - 1998



Fonte: Tabela 02

GRÁFICO EM SETORES

- É a representação através de um círculo, por meio de setores.
- Muito utilizado quando pretendemos comparar cada valor da série com o total - proporção.
- Forma de cálculo:

$$\begin{array}{lcl} \text{Total} & \underline{\qquad\qquad} & 360^{o} \\ \text{parte} & \underline{\qquad\qquad} & x^{o} \end{array}$$

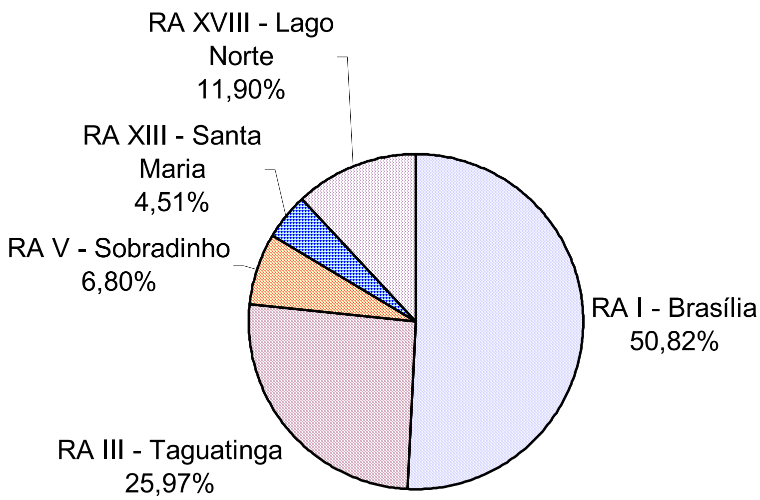
Tabela 03. Efetivo (valores absoluto e relativo) do CBMDF em Cinco Regiões Administrativas do DF - 1998
FONTE: Banco de Dados do Distrito Federal – 1998

Região Administrativa	Efetivo	
	Absoluto	Relativo (%)
RA I - Brasília	867	50,82
RA III - Taguatinga	443	25,97
RA V - Sobradinho	116	6,80
RA XIII - Santa Maria	77	4,51
RA XVIII - Lago Norte	203	11,90
Total	1.706	100,00

NOTAS: Os efetivos especializados (emergência médica, incêndio florestal e guarda e segurança) estão alocados nas regiões administrativas.

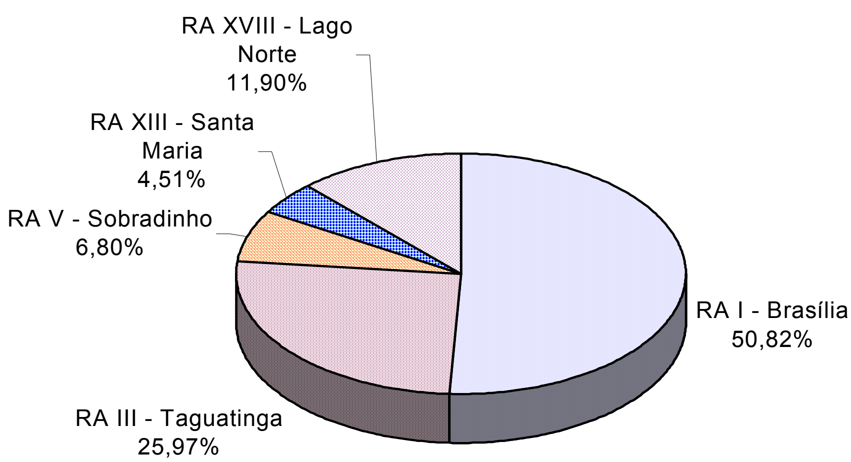
	Efetivo		x^{o}
RA I - Brasília	867	$\longrightarrow$	183,0
RA III - Taguatinga	443	$\longrightarrow$	93,5
RA V - Sobradinho	116	$\longrightarrow$	24,5
RA XIII - Santa Maria	77	$\longrightarrow$	16,2
RA XVIII - Lago Norte	203	$\longrightarrow$	42,8
Total	1.706		360,0

Gráfico 03.a. Comparativo (percentual) do Efetivo do CBMDF em Cinco Regiões Administrativas do DF – 1998



FONTE: Tabela 03

Gráfico 03.b. Comparativo (percentual) do Efetivo do CBMDF em Cinco Regiões Administrativas do DF – 1998



FONTE: Tabela 03

GRÁFICO EM CURVAS / LINHAS

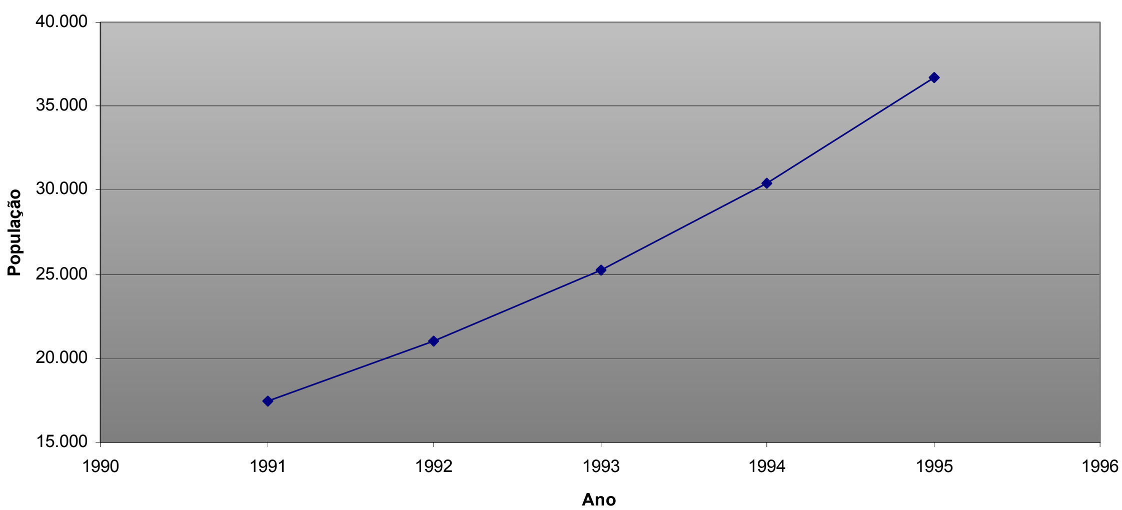
- Muito utilizado para representar dados ***temporais***.

Tabela 04. População da RA XIV – São Sebastião – 1991 a 1995

Ano	População
1991	17.399
1992	20.971
1993	25.271
1994	30.457
1995	36.703

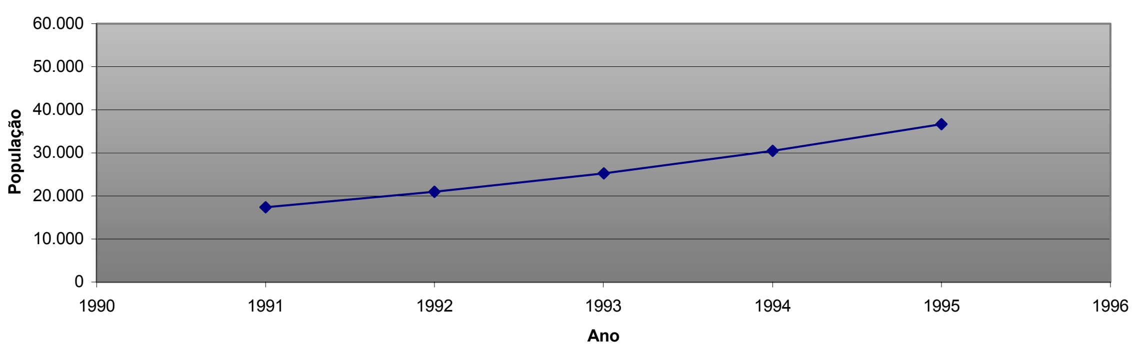
FONTE: Censo Demográfico de 1991 – IBGE
Estimativas para 1992 a 1995 - CODEPLAN

Gráfico 04. População da RA XIV – São Sebastião – 1991 a 1995



FONTE: Tabela 04

Gráfico 05. População da RA XIV – São Sebastião – 1991 a 1995

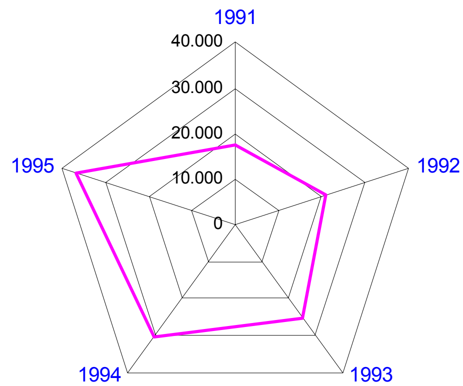


FONTE: Tabela 04

GRÁFICO POLAR / RADAR

- ❖ Representação por meio de um polígono
- ❖ Geralmente presta-se para apresentação de séries temporais

Gráfico 05. População da RA XIV – São Sebastião – 1991 a 1995



FONTE: Tabela 04

EXERCÍCIOS

1. Assinale a afirmativa verdadeira:
    a) Um gráfico de barras ou colunas é aquele em que os retângulos que o compõem estão dispostos horizontalmente.
    b) Um gráfico de barras ou colunas é aquele em que os retângulos que o compõem estão dispostos verticalmente.
    c) Um gráfico de barras é aquele em que os retângulos que o compõem estão dispostos verticalmente e um gráfico de colunas, horizontalmente.
    d) Um gráfico de barras é aquele em que os retângulos que o compõem estão dispostos horizontalmente e um gráfico de colunas, verticalmente.
    e) Todas as alternativas anteriores são falsas.

2. O gráfico mais comumente utilizado quando se deseja evidenciar a participação de um dado em relação ao total é denominado:
    a) Gráfico em barras;
    b) Gráficos em colunas;
    c) Gráfico em setores;
    d) Gráfico pictórico ou pictograma;
    e) Gráfico decorativo.

3. Uma representação gráfica comumente encontrada em jornais e revistas que inclui figuras de modo a torná-las mais atraente é denominada:
    a) Gráfico em barras;
    b) Gráficos em colunas;
    c) Gráfico em setores;
    d) Gráfico pictórico ou pictograma;
    e) Gráfico decorativo.

4. A tabela abaixo mostra o consumo de determinada bebida durante um baile de carnaval:

Bebida	Consumo (l)
Vinho	100
Suco de Frutas	200
Água Mineral	400
Refrigerante	700
Cerveja	1600

Foi construído um gráfico em setores para melhor representar o fenômeno acima.
    a) Qual o ângulo do setor correspondente ao vinho?
        i) 6°
        ii) 10°
        iii) 12°
        iv) 24°
        v) 100°

    b) Qual o ângulo do setor correspondente ao suco de frutas?
        i) 12°
        ii) 20°
        iii) 24°
        iv) 48°
        v) 200°
    c) Qual o ângulo do setor correspondente à água mineral?

i) 24°
ii) 40°
iii) 48°
iv) 84°
v) 100°

d) Qual o ângulo do setor correspondente aos refrigerantes?
i) 42°
ii) 70°
iii) 84°
iv) 192°
v) 700°

e) Qual o ângulo do setor correspondente às cervejas?
i) 12°
ii) 96°
iii) 160°
iv) 192°
v) 1600°

UNIDADE IV - DISTRIBUIÇÃO DE FREQÜÊNCIA

REPRESENTAÇÃO DA AMOSTRA:

Podemos observar que a estatística tem como objetivo encontrar leis de comportamento para todo o conjunto, por meio da sintetização dos dados numéricos, sob a forma de tabelas, gráficos e medidas.

PROCEDIMENTO COMUM PARA A REPRESENTAÇÃO DAS DISTRIBUIÇÕES DE FREQÜÊNCIA (MANEIRA DE SUMARIZAR OS DADOS)

1) DADOS BRUTOS: O conjunto dos dados numéricos obtidos após a crítica dos valores coletados constitui-se nos dados brutos. Assim:

24	23	22	28	35	21	23	23	33	34
24	21	25	36	26	22	30	32	25	26
33	34	21	31	25	31	26	25	35	33

4) ROL: É o arranjo dos dados brutos em ordem de freqüências crescente ou decrescente: Assim:

21	21	21	22	22	23	23	23	24	24
25	25	25	25	26	26	26	28	30	31
31	32	33	33	33	34	34	35	35	36

3) AMPLITUDE TOTAL OU RANGE "R" : É a diferença entre o maior e o menor valor observado.

No exemplo: R = 36 - 21 = 15

4) FREQÜÊNCIA ABSOLUTA (F_i): É o número de vezes que o elemento aparece na amostra, ou o número de elementos pertencentes a uma classe.
No exemplo $F_{(21)} = 3$.

5) DISTRIBUIÇÃO DE FREQÜÊNCIA: É o arranjo dos valores e suas respectivas freqüências. Assim, a distribuição de freqüência para o exemplo será:

X_i	F_i
21	3
22	2
23	3
24	2
25	4
26	3
28	1
30	1
31	2
32	1
33	3
34	2
35	2
36	1
$\sum$	30

Para a variável contínua:
Seja X_i peso de 100 indivíduos:

CLASSE	F_i
45 |- 55	15
55 |- 65	30
65 |- 75	35
75 |- 85	15
85 |- 95	5
$\sum$	100

6) NUMERO DE CLASSES (K): Não há fórmula exata para o número de classes (arredondar para o inteiro mais próximo). Soluções:

- $K = \begin{cases} 5, \text{se } n < 25 \\ \sqrt{n}, \text{se } n \geq 25 \end{cases}$
- Fórmula de Sturges: K= 1 + 3,32 log(n)

Onde: n = tamanho da amostra.

EXEMPLO:

Considere o exemplo apresentado no ROL:

$$K = 1 + 3{,}32 \cdot \log(30) \Rightarrow K = 5{,}9 \Rightarrow K = 6$$

Portanto, a tabela irá conter 6 classes.

7) AMPLITUDE DA CLASSE (h): $h = \dfrac{R}{K}$ (aproximar para o maior inteiro).

EXEMPLO:

Considere novamente o exemplo apresentado no ROL:

$$h = \frac{15}{6} \Rightarrow h = 2{,}5 \Rightarrow h = 3$$

8) LIMITE DE CLASSES: Representado por

10 |-| 12: valores entre 10 e 12;

10 -| 12: valores de 10 a 12, excluindo o 10;

10 |- 12: valores de 10 a 12, excluindo o 12.

Obs.: Neste curso iremos utilizar a última representação.

EXEMPLO:

Considere o exemplo apresentado no ROL:

Classe	F_i
21 |- 24	8
24 |- 27	9
27 |- 30	1
30 |- 33	4
33 |- 36	7
36 |- 39	1
TOTAL	30

9) PONTO MÉDIO DA CLASSE (x_i): É a média aritmética entre o limite superior (L_i) e o inferior da classe (l_i).

$$x_i = \frac{l_i + L_i}{2}$$

EXEMPLO:

Da tabela acima:

Classe	F_i	$\mathbf{x_i}$
21 |- 24	8	**22,5**
24 |- 27	9	**25,5**
27 |- 30	1	**28,5**
30 |- 33	4	**31,5**
33 |- 36	7	**34,5**
36 |- 39	1	**37,5**
TOTAL	30	**-**

10) FREQÜÊNCIA ABSOLUTA ACUMULADA (F_{ac}): É a soma das freqüências dos valores inferiores ou iguais ao valor dado.

Exemplo:

Classe	F_i	x_i	$\mathbf{F_{ac}}$
21 |- 24	8	22,5	**8**
24 |- 27	9	25,5	**17**
27 |- 30	1	28,5	**18**
30 |- 33	4	31,5	**22**
33 |- 36	7	34,5	**29**
36 |- 39	1	37,5	**30**
TOTAL	30	-	**-**

11) FREQÜÊNCIA RELATIVA SIMPLES (f_i): A freqüência relativa de um valor é dada por, $f_i = \frac{F_i}{\sum F_i}$, ou será a percentagem daquele valor na amostra caso multiplique por 100.

Classe	F_i	x_i	F_{ac}	f_i
21 |- 24	8	22,5	8	**0,267**
24 |- 27	9	25,5	17	**0,300**
27 |- 30	1	28,5	18	**0,033**
30 |- 33	4	31,5	22	**0,133**
33 |- 36	7	34,5	29	**0,233**
36 |- 39	1	37,5	30	**0,033**
TOTAL	30	-	-	**1,000**

Exemplo:

12) FREQÜÊNCIA RELATIVA ACUMULADA (f_{ac}): É a soma das freqüências relativas dos valores inferiores ou iguais ao valor dado.

Exemplo:

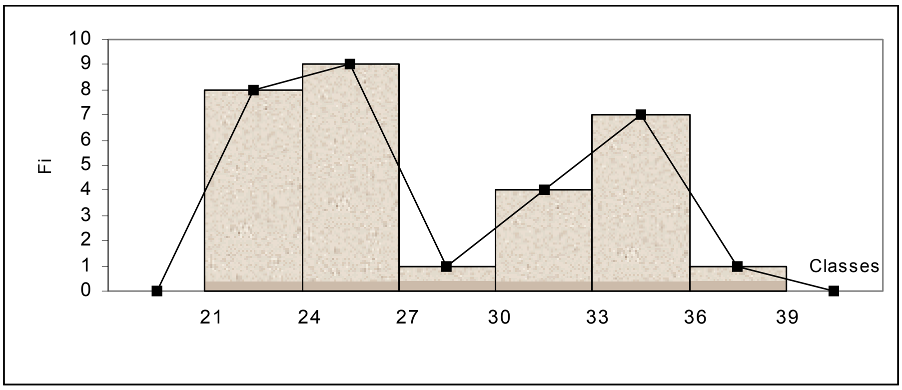
Classe	F_i	x_i	F_{ac}	f_i	f_{ac}
21 |- 24	8	22,5	8	0,267	**0,267**
24 |- 27	9	25,5	17	0,300	**0,567**
27 |- 30	1	28,5	18	0,033	**0,600**
30 |- 33	4	31,5	22	0,133	**0,733**
33 |- 36	7	34,5	29	0,233	**0,966**
36 |- 39	1	37,5	30	0,033	**1,000**
TOTAL	30	-	-	1,000	**-**

13) HISTOGRAMA: É a representação gráfica de uma distribuição de FREQÜÊNCIA por meio de retângulos justapostos (veja exemplo a seguir).

14) POLÍGONO DE FREQÜÊNCIA: É a representação gráfica de uma distribuição por meio de um polígono.

Exemplo:

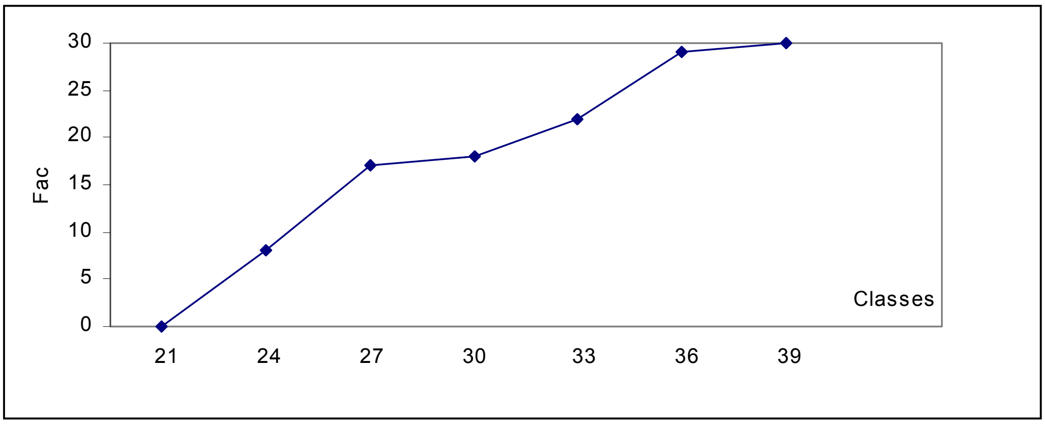
HISTOGRAMA E POLÍGONO DE FREQUÊNCIA SIMPLES DA TABELA ACIMA



15) POLÍGONO DE FREQÜÊNCIA ACUMULADA:

Exemplo:

POLÍGONO DE FREQUÊNCIA ACUMULADA DA TABELA ACIMA



EXERCÍCIOS

1. Um dado foi lançado 50 vezes e foram registrados os seguintes resultados

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

Construa uma distribuição de freqüência sem intervalo de classe e determine:

a) A Amplitude Total
   i) 5
   ii) 6
   iii) 7
   iv) 10
   v) 50

b) A freqüência total
   i) 5
   ii) 6
   iii) 7
   iv) 10
   v) 50

c) A freqüência simples absoluta do primeiro elemento:
   i) 10%
   ii) 20%
   iii) 1
   iv) 10
   v) 20

d) A freqüência simples relativa do primeiro elemento:
   i) 10%
   ii) 20%
   iii) 1
   iv) 10
   v) 20

e) A freqüência acumulada do primeiro elemento:
   i) 10%
   ii) 20%
   iii) 1
   iv) 10
   v) 20

f) A freqüência acumulada relativa do primeiro elemento:
   i) 10%
   ii) 20%
   iii) 1
   iv) 10
   v) 20

g) A freqüência simples absoluta do segundo elemento:
   i) 19
   ii) 9
   iii) 2
   iv) 38%

v) 18%

h) A freqüência simples relativa do quinto elemento:
   i) 12%
   ii) 84%
   iii) 5
   iv) 6
   v) 42

i) A freqüência acumulada relativa do sexto elemento:
   i) 50
   ii) 8
   iii) 6
   iv) 100%
   v) 16%

3. Dado o rol de medidas das alturas (dadas em cm) de uma amostra de 100 indivíduos de uma faculdade:

151	152	154	155	158	159	159	160	161	161
161	162	163	163	163	164	165	165	165	166
166	166	166	167	167	167	167	167	168	168
168	168	168	168	168	168	168	168	169	169
169	169	169	169	169	170	170	170	170	170
170	170	171	171	171	171	172	172	172	173
173	173	174	174	174	175	175	175	175	176
176	176	176	177	177	177	177	178	178	178
179	179	180	180	180	180	181	181	181	182
182	182	183	184	185	186	187	188	190	190

Calcule:

a) A amplitude amostral;

b) O número de classes;

c) A amplitude de classes;

d) Os limites de classes;

e) As freqüências absolutas das classes;

f) As freqüências relativas;

g) Os pontos médios das classes;

h) As freqüências acumuladas;

i) O histograma e o polígono de freqüência;

j) O polígono de freqüência acumulada;

k) Faça um breve comentário sobre os valores das alturas desta amostra através da distribuição de frequência.

4. Os dados seguintes representam 20 observações relativas ao índice pluviométrico em determinado município do Estado:

Milímetros de chuva

144	152	159	160
160	151	157	146
154	145	151	150
142	146	142	141
141	150	143	158

a) Determinar o número de classes pela regra de Sturges;
b) Construir a tabela de freqüências absolutas simples;

c) Determinar as freqüências absolutas acumuladas;
d) Determinar as freqüências simples relativas;

5. Considere a seguinte distribuição de frequência correspondente aos diferentes preços de um determinado produto em vinte lojas pesquisadas.

Preços	No. De lojas
50	2
51	5
52	6
53	6
54	1
Total	20

a) Quantas lojas apresentaram um preço de R$52,00?
b) Construa uma tabela de freqüências simples relativas.
c) Construa uma tabela de freqüências absolutas acumuladas.
d) Quantas lojas apresentaram um preço de até R$52,00 (inclusive)?
e) Qual o percentual de lojas com preço maior de que R$51,00 e menor de que R$54,00?

6. O quadro seguinte representa as alturas (em cm) de 40 alunos de uma classe.

162	163	148	166	169	154	170	166
164	165	159	175	155	163	171	172
170	157	176	157	157	165	158	158
160	158	163	165	164	178	150	168
166	169	152	170	172	165	162	164

a) Calcular a amplitude total.
b) Admitindo-se 6 classes, qual a amplitude do intervalo de classe?
c) Construir uma tabela de frequência das alturas dos alunos.
d) Determinar os pontos médios das classes.

7. Vinte alunos foram submetidos a um teste de aproveitamento cujos resultados são.

26	28	24	13	18
18	25	18	25	24
20	21	15	28	17
27	22	13	19	28

Pede-se agrupar tais resultados em uma distribuição de freqüências:

UNIDADE V - MEDIDAS DE POSIÇÃO E SEPARATRIZES

MEDIDAS DE POSIÇÃO

As medidas de posição, também chamada de medidas de tendência central, possuem três formas diferentes para três situações distintas:

MÉDIA ARITMÉTICA

Existem duas médias:

- POPULACIONAL, representada letra grega μ
- AMOSTRAL, representada por $\overline{x}$

1ª SITUAÇÃO: Dados não agrupados

Sejam os elementos x_1, x_2, x_3,...,x_n de uma amostra, portanto "n" valores da variável X. A média aritmética da variável aleatória de X é definida por,

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n} \quad \text{ou simplesmente,} \quad \overline{x} = \frac{\sum x}{n}$$

Onde n é o número de elementos do conjunto.

Exemplo:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 7, 8, 10 e 11. Determinar a média aritmética simples deste conjunto de dados.

$$\overline{x} = \frac{3+7+8+10+11}{5} = \frac{39}{5} = 7{,}8$$

Interpretação: o tempo médio de serviço deste grupo de funcionários é de 7,8 anos.

2ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por valores simples

Quando os dados estiverem agrupados numa distribuição de freqüência usaremos a média aritmética dos valores x_1, x_2, x_3,...,x_n, ponderados pelas respectivas freqüências absolutas: F_1, F_2, F_3, ... , F_n. Assim

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i F_i}{n}$$

Exemplo:

Em um determinado dia foi registrado o número de veículos negociados por uma amostra de 10 vendedores de uma agência de automóveis obtendo a seguinte tabela:

veículos negociados (x_i)	número de vendedores (F_i)	**x_i F_i**
1	1	**1**
2	3	**6**
3	5	**15**
4	1	**4**
TOTAL	10	**26**

Portanto:

$$\overline{x} = \frac{26}{10} = 2,6$$

Interpretação: em média, cada vendedor negociou 2,6 veículos.

3ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por classes

Quando os dados estiverem agrupados numa distribuição de freqüência usaremos a média aritmética dos pontos médios x_1, x_2, x_3,...,x_n de cada classe, ponderados pelas respectivas freqüências absolutas: F_1, F_2, F_3, ... , F_n. Desta forma, o cálculo da média passa a ser igual ao da 2ª situação. Assim

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i F_i}{n}$$

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada disciplina:

ESCORES	ALUNOS (F_i)	x_i	x_i F_i
35 |- 45	5	40	200
45 |- 55	12	50	600
55 |- 65	18	60	1.080
65 |- 75	14	70	980
75 |- 85	6	80	480
85 |- 95	3	90	270
TOTAL	58	-	3.610

Portanto,

$$\overline{x} = \frac{3610}{58} = 62,24$$

Interpretação: o desempenho médio deste grupo de alunos foi de 62,24 pontos nesta disciplina.

MODA - Mo

Dentre as principais medidas de posição, destaca-se a moda. É o valor *mais freqüente* da distribuição.

↳ **1ª SITUAÇÃO: Dados não agrupados**

Sejam os elementos x_1, x_2, x_3,...,x_n de uma amostra, o valor da moda para este tipo de conjunto de dados é simplesmente o valor com maior frequência.

Exemplo 1:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 7, 8, 8 e 11. Determinar a moda deste conjunto de dados.

$$\text{Mo} = 8 \Rightarrow \text{Distribuição unimodal ou modal}$$

Interpretação: o tempo de serviço com maior frequência é de 8 anos.

Exemplo 2:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 3, 7, 8, 8 e 11. Determinar a moda deste conjunto de dados.

$$\left.\begin{array}{l}\text{Mo} = 3\\ \text{Mo} = 8\end{array}\right\} \Rightarrow \text{Distribuição bimodal}$$

Interpretação: os tempos de serviço com maior frequência foram de 3 e 8 anos.

Exemplo 3:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 7, 8, 10 e 11. Determinar a moda deste conjunto de dados.

$$\text{não existe Mo} \Rightarrow \text{Distribuição amodal}$$

Interpretação: não existe o tempo de serviço com maior frequência.

↳ **2ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por valores simples**

Para este tipo de distribuição, a identificação da moda é facilitada pela simples observação do elemento que apresenta maior freqüência. Assim, para a distribuição.

Exemplo:

Em um determinado dia foi registrado o número de veículos negociados por uma amostra de 10 vendedores de uma agência de automóveis obtendo a seguinte tabela:

veículos negociados (x_i)	número de vendedores (F_i)
1	1
2	3
3	**5**
4	1
TOTAL	10

Portanto, se a maior frequência é $F_i = 5$, logo Mo = 3.

Interpretação: A quantidade de veículos comercializados no dia com maior frequência foi de três veículos.

➭ 3ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por classes

Para dados agrupados em classes, temos diversas fórmulas para o cálculo da moda. A utilizada será:

Fórmula de Czuber

Procedimento:

a) Identifica-se a classe modal (aquela que possuir maior freqüência) – CLASSE(Mo).

b) Utiliza-se a fórmula:

$$Mo = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} \cdot h$$

em que:

$$l_i = \text{limite inferior da classe modal}$$
$$\Delta_1 = F_i - F_{i,ant}$$
$$\Delta_2 = F_i - F_{i,post}$$
$$h = \text{amplitude da classe modal}$$

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada disciplina:

ESCORES	ALUNOS F_i
35 |- 45	5
45 |- 55	12
55 |- 65	**18**
65 |- 75	14
75 |- 85	6
85 |- 95	3
TOTAL	58

$$CLASSE(Mo) \Rightarrow 55 \mid -65$$

$$Mo = 55 + \frac{6}{6+4} \cdot 10 = 55 + 6 \Rightarrow Mo = 61$$

$$\textit{onde}:$$

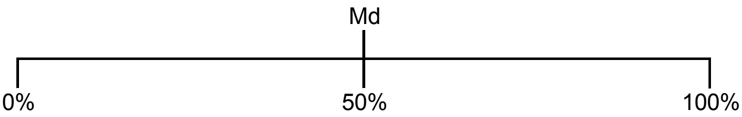
$$\Delta_1 = 18 - 12 = 6$$

$$\Delta_2 = 18 - 14 = 4$$

Interpretação: O escore com maior frequência entre o grupo de 58 alunos foi de 61 pontos.

MEDIANA - Md

Construído o ROL, o valor da mediana é o elemento que ocupa a posição central, ou seja, é o elemento que divide a distribuição em 50% de cada lado:



↳ **1ª SITUAÇÃO: Dados não agrupados**

Sejam os elementos x_1, x_2, x_3,...,x_n de uma amostra, portanto "n" valores da variável X. A mediana da variável aleatória de X é definida por,

$$\text{se n} = \begin{cases} \text{par, então o valor da mediana será a média das duas observações adjacentes à posição } \dfrac{n+1}{2} \\ \text{ímpar, então o valor da mediana será o valor localizado na posição } \dfrac{n+1}{2} \end{cases}$$

Exemplo 1:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 7, 8, 10 e 11. Determinar a mediana deste conjunto de dados.

Como $n = 5$, então o valor da mediana estará localizado na posição $\frac{5+1}{2} = 3$. Portanto,

$$\text{Md} = 8$$

<u>Interpretação</u>: 50% dos funcionários possuem até oito anos de tempo de serviço, ou, 50% dos funcionários possuem no mínimo oito anos de tempo de serviço.

Exemplo 2:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 7, 8, 10, 11 e 13. Determinar a mediana deste conjunto de dados.

Como $n = 6$, então o valor da mediana estará localizado na posição $\frac{6+1}{2} = 3{,}5$. Portanto,

$$Md = \frac{8+10}{2} = 9$$

<u>Interpretação</u>: 50% dos funcionários possuem até nove anos de tempo de serviço, ou, 50% dos funcionários possuem no mínimo nove anos de tempo de serviço.

↳ **2ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por valores simples**

Quando os dados estiverem agrupados numa distribuição de freqüência identificaremos a mediana dos valores x_1, x_2, x_3,...,x_n pela posição da mediana $\text{POS}(\text{Md}) = \frac{n}{2}$ através da frequência absoluta acumulada - F_{ac},

Exemplo:

Em um determinado dia foi registrado o número de veículos negociados por uma amostra de 10 vendedores de uma agência de automóveis obtendo a seguinte tabela:

veículos negociados (x_i)	número de vendedores (F_i)	F_{ac}
1	1	**1**
2	3	**4**
3	5	**9**
4	1	**10**
TOTAL	10	**-**

Portanto:

$$POS(Md) = \frac{10}{2} = 5 \Rightarrow Md = 3$$

Interpretação: 50% dos vendedores comercializaram no máximo três veículos, ou então, metade dos vendedores comercializou pelo menos três veículos.

3ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por classes

Procedimento:

1. Calcula-se a posição da mediana: $POS(Md) = \frac{n}{2}$
2. Pela F_{ac} identifica-se a classe que contém o valor da mediana - CLASSE(Md)
3. Utiliza-se a fórmula: $Md = l_i + \frac{POS(Md) - F_{ac,ant}}{F_i} \cdot h$

Onde:

l_i = Limite inferior da classe mediana

n = Tamanho da amostra ou número de elementos

$F_{ac,ant}$ = Frequência acumulada anterior à classe mediana

h = Amplitude da classe mediana

F_i = Freqüência absoluta simples da classe mediana

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada disciplina:

ESCORES	ALUNOS (F_i)	F_{ac}
35 |- 45	5	**5**
45 |- 55	12	**17**
55 |- 65	18	**35**
65 |- 75	14	**49**
75 |- 85	6	**55**
85 |- 95	3	**58**
TOTAL	58	**-**

Portanto,

$$1.\ \text{POS(Md)} = \frac{58}{2} = 29$$

$$2.\ \text{CLASSE(Md)} = 55 \mid -65$$

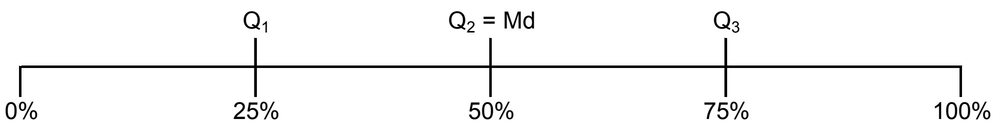
$$3.\ \text{Md} = 55 + \frac{29-17}{18} \cdot 10 = 55 + 6{,}67 \Rightarrow \text{Md} = 61{,}67$$

<u>Interpretação</u>: 50% dos alunos obtiveram escore máximo de 61,67 pontos, ou então, metade dos alunos obtiveram escore maior que 61,67 pontos..

SEPARATRIZES

QUARTIS

Os quartis dividem um conjunto de dados em quatro partes iguais.



Assim:

Onde:

Q_1 = 1° quartil, deixa 25% dos elementos.

Q_2 = 2° quartil, coincide com a mediana, deixa 50% dos elementos.

Q_3 = 3° quartil, deixa 75% dos elementos.

Procedimento:

1. Calcula-se a posição do quartil: $\text{POS}(Q_i) = \frac{n}{4} \cdot i$

   $\text{onde}: i = 1,2,3$

2. Pela F_{ac} identifica-se a classe que contém o valor do quartil - CLASSE(Q_i)

3. Utiliza-se a fórmula: $$Q_i = l_i + \frac{POS(Q_i) - F_{ac,ant}}{F_i} \cdot h$$

onde:

l_i = Limite inferior da classe quartílica

n = Tamanho da amostra ou número de elementos

$F_{ac,ant}$ = Frequência acumulada anterior à classe quartílica

h = Amplitude da classe quartílica

F_i = Freqüência absoluta simples da classe quartílica

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada

ESCORES	ALUNOS (F_i)	F_{ac}
35 |- 45	5	**5**
45 |- 55	12	**17**
55 |- 65	18	**35**
65 |- 75	14	**49**
75 |- 85	6	**55**
85 |- 95	3	**58**
TOTAL	58	**-**

disciplina. Calcule o primeiro e o terceiro quartil.

Portanto,

$$1.\ POS(Q_1) = \frac{58}{4} \cdot 1 = 14{,}5$$
$$2.\ CLASSE(Q_1) = 45 \mid -55$$
$$3.\ Q_1 = 45 + \frac{14{,}5 - 5}{12} \cdot 10 = 45 + 7{,}92 \Rightarrow Q_1 = 52{,}92$$

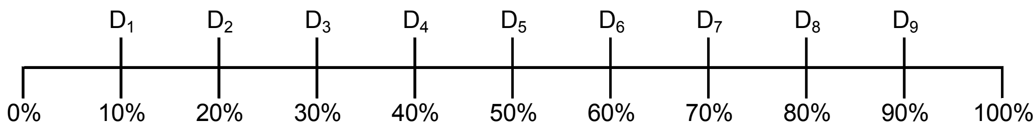
Interpretação: 25% dos alunos obtiveram escore máximo de 52,92 pontos, ou então, 75% dos alunos obtiveram escore maior que 52,92 pontos.

$$1.\ POS(Q_3) = \frac{58}{4} \cdot 3 = 43{,}5$$
$$2.\ CLASSE(Q_1) = 65 \mid -75$$
$$3.\ Q_3 = 65 + \frac{43{,}5 - 35}{14} \cdot 10 = 65 + 6{,}07 \Rightarrow Q_3 = 71{,}07$$

Interpretação: 75% dos alunos obtiveram escore menor que 71,07 pontos, ou então, 25% dos alunos obtiveram escore de pelo menos 71,07 pontos.

DECIS

São valores que divide a série em dez partes.



Procedimento:

1. Calcula-se a posição da medida: $\mathrm{POS}(D_i) = \frac{n}{10} \cdot i$

   $\text{onde}: i = 1,2,3,4,5,6,7,8,9$

2. Pela F_{ac} identifica-se a classe que contém o valor do decil - CLASSE(D_i)
3. Utiliza-se a fórmula: $D_i = l_i + \frac{\mathrm{POS}(D_i) - F_{ac,ant}}{F_i} \cdot h$

Onde:

l_i = Limite inferior da classe do decil

n = Tamanho da amostra ou número de elementos

$F_{ac,ant}$ = Frequência acumulada anterior à classe do decil

h = Amplitude da classe do decil

F_i = Freqüência absoluta simples da classe do decil

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada disciplina. Calcule o sexto decil.

ESCORES	ALUNOS (F_i)	F_{ac}
35 |- 45	5	**5**
45 |- 55	12	**17**
55 |- 65	18	**35**
65 |- 75	14	**49**
75 |- 85	6	**55**
85 |- 95	3	**58**
TOTAL	58	**-**

Portanto,

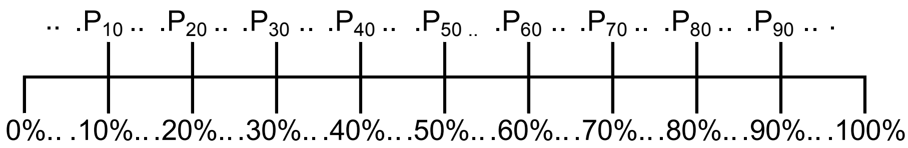
$$1.\ \text{POS}(D_6) = \frac{58}{10} \cdot 6 = 34{,}8$$

$$2.\ \text{CLASSE}(D_6) = 55 \mid\!-65$$

$$3.\ D_6 = 55 + \frac{34{,}8 - 17}{18} \cdot 10 = 55 + 9{,}89 \Rightarrow D_6 = 64{,}89$$

Interpretação: 60% dos alunos obtiveram escore inferior a 64,89 pontos, ou então, 40% dos alunos obtiveram escore mínimo de 64,89 pontos.

PERCENTIS



São as medidas que dividem a amostra em 100 partes iguais. A fórmula será:

Procedimento:

1. Calcula-se a posição da medida: $\text{POS}(P_i) = \frac{n}{100} \cdot i$

   $\text{onde}: i = 1,2,3,...,98,99$

2. Pela F_{ac} identifica-se a classe que contém o valor do percentil - CLASSE(P_i)
3. Utiliza-se a fórmula: $P_i = l_i + \frac{\text{POS}(P_i) - F_{ac,ant}}{F_i} \cdot h$

onde:

l_i = Limite inferior da classe do percentil

n = Tamanho da amostra ou número de elementos

$F_{ac,ant}$ = Frequência acumulada anterior à classe do percentil

h = Amplitude da classe do percentil

F_i = Freqüência absoluta simples da classe do percentil

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada disciplina. Calcule o percentil de ordem 23.

ESCORES	ALUNOS (F_i)	F_{ac}
35 |- 45	5	**5**
45 |- 55	12	**17**
55 |- 65	18	**35**
65 |- 75	14	**49**
75 |- 85	6	**55**
85 |- 95	3	**58**
TOTAL	58	**-**

Portanto,

$$
\begin{aligned}
&1.\,\mathrm{POS}(P_{23}) = \frac{58}{100}\cdot 23 = 13{,}34\\
&2.\,\mathrm{CLASSE}(P_{23}) = 45 \,|\!-55\\
&3.\,P_{23} = 45 + \frac{13{,}34 - 5}{12}\cdot 10 = 45 + 6{,}95 \Rightarrow P_{23} = 51{,}95
\end{aligned}
$$

Interpretação: 23% dos alunos com os menores escores obtiveram pontuação inferior a 51,95 pontos, ou então, 77% dos alunos obtiveram escore maior que 51,95 pontos.

EXERCÍCIOS

1. Dado o rol do número de erros de impressão da primeira página de um jornal durante 50 dias, obteve-se os seguintes resultados:

5	5	5	6	6	6	7	7	7	7
7	8	8	8	8	8	8	8	9	9
10	10	10	10	10	11	11	11	11	12
12	12	12	12	12	12	12	12	13	14
14	14	14	14	14	14	15	16	19	22

a) Complete a tabela de distribuição de frequência:

Classe	Fi	xi	Fac	f_i
05 |- 08				
08 |- 11				
11 |- 14				
14 |- 17				
17 |- 20				
20 |- 23				
Total		-	-	

Segundo nos mostra a tabela acima responda:

i) Qual a amplitude total (r)?

ii) Qual o valor de k (número de classe)?

iii) Qual o intervalo de cada classe (h)?

2. Complete a tabela a seguir:

Classes	f	P.M.	Fi	fr
				0,02
	12			
62 - 65				0,06
		66,5	84	
			126	
	36			
			225	
				0,15
			300	
Total		-	-	

3. Considere a seguinte tabela:

classes	Fi
2,75 |- 2,80	2
2,80 |- 2,85	3
2,85 |- 2,90	10
2,90 |- 2,95	11
2,95 |- 3,00	24
3,00 |- 3,05	14
3,05 |- 3,10	9
3,10 |- 3,15	8
3,15 |- 3,20	6
3,20 |- 3,25	3
Total	90

Identificar os seguintes elementos da tabela:

a) Frequência simples absoluta da quinta classe.

b) Frequência total.

c) Limite inferior da sexta classe.
d) Limite superior da quarta classe.
e) Amplitude do intervalo de classe.
f) Amplitude total.
g) Ponto médio da terceira classe.
h) Número total de classe.
i) Frequência absoluta acumulada além da sexta classe.
j) Porcentagem de valores iguais ou maiores que 3,20.

4. Responda as questões abaixo:

I) Média, Mediana e Moda são medidas de :
a) ( ) Dispersão b) ( ) posição
c) ( ) assimetria d) ( ) curtose

II) Na série 10, 20, 40, 50, 70, 80 a mediana será:
a) ( ) 30 b) ( ) 35
c) ( ) 40 d) ( ) 45

III) 50% dos dados da distribuição situa-se:
a) ( ) abaixo da média c) ( ) abaixo da moda
b) ( ) acima da mediana d) ( ) acima da média

8. Calcule para cada caso abaixo a respectiva média.
a) 7, 8, 9, 12, 14

b)

Xi	3	4	7	8	12
Fi	2	5	8	4	3

c)

Classes	68 - 72	72 - 76	76 - 80	80 - 84
Fi	8	20	35	40

9. Calcule o valor da mediana.
a) 82, 86, 88, 84, 91, 93

b)

Xi	73	75	77	79	81
Fi	2	10	12	5	2

c)

Classes	1 - 3	3 - 5	5 - 7	7 - 9	9 - 11	11 - 13
Fi	3	5	8	6	4	3

10. Calcule a moda
a) 3, 4, 7, 7, 7, 8, 9, 10

b)

Xi	2,5	3,5	4,5	6,5
Fi	7	17	10	5

c)

Classes	10 - 20	20 - 30	30 - 40	40 - 50
Fi	7	19	28	32

11. Para a distribuição abaixo calcular D_2, P_4 Q_3.

a)

Classes	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70
Fi	3	8	18	22	24

UNIDADE VI - MEDIDAS DE DISPERSÃO

MEDIDA DE DISPERSÃO

As medidas de dispersão indicam se os valores estão relativamente próximos um dos outros, ou separados em torno de uma medida de posição: a média. Consideraremos quatro medidas de dispersão: Desvio-médio, Variância, Desvio Padrão e Coeficiente de Variação.

DESVIO-MÉDIO

O desvio-médio analisa a média dos desvios em torno da média.

➾ **1ª SITUAÇÃO: Dados não agrupados**

Sejam os elementos $x_1, x_2, x_3, \ldots, x_n$ de uma amostra, portanto "n" valores da variável X, com média igual a $\overline{x}$. O desvio-médio da variável aleatória de X é,

$$DM = \frac{\sum |x_i - \overline{x}|}{n}$$

onde n é o número de elementos do conjunto.

Exemplo:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 7, 8, 10 e 11. Determinar o desvio-médio deste conjunto de dados.

$$\text{como } \overline{x} = 7{,}8$$

$$\text{então } DM = \frac{|3 - 7{,}8| + |7 - 7{,}8| + |8 - 7{,}8| + |10 - 7{,}8| + |11 - 7{,}8|}{5} = \frac{11{,}2}{5} \Rightarrow DM = 2{,}24$$

<u>Interpretação</u>: em média, o tempo de serviço deste grupo de funcionários se desvia em 2,24 anos em torno dos 7,8 anos de tempo médio de serviço.

➾ **2ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por valores simples**

Quando os dados estiverem agrupados numa distribuição de freqüência usaremos o desvio-médio dos valores $x_1, x_2, x_3, \ldots, x_n$, ponderados pelas respectivas freqüências absolutas: $F_1, F_2, F_3, \ldots, F_n$, como no cálculo da média aritmética. Assim

$$DM = \frac{\sum |x_i - \overline{x}| \cdot F_i}{n}$$

Exemplo:

Em um determinado dia foi registrado o número de veículos negociados por uma amostra de 10 vendedores de uma agência de automóveis como mostra a tabela abaixo. O cálculo do desvio-médio será:

veículos negociados (x_i)	número de vendedores (F_i)	**|xi-média|**	**|xi-média|*Fi**
1	1	**1,60**	**1,60**
2	3	**0,60**	**1,80**
3	5	**0,40**	**2,00**
4	1	**1,40**	**1,40**
TOTAL	10	**4,00**	**6,80**

$$\text{como } \bar{x} = 2{,}6$$

$$\text{então DM} = \frac{6{,}8}{10} = 0{,}68$$

<u>Interpretação</u>: em média, a quantidade de veículos negociada de cada vendedor possuiu uma distância de 0,68 em torno dos 2,6 veículos comercializados em média por vendedor.

➪ **3ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por classes**

Quando os dados estiverem agrupados numa distribuição de freqüência usaremos o desvio-médio dos pontos médios x_1, x_2, x_3,...,x_n de cada classe, ponderados pelas respectivas freqüências absolutas: F_1, F_2, F_3, ... , F_n. Desta forma, o cálculo do desvio-médio passa a ser igual ao da 2ª situação. Assim

$$DM = \frac{\sum |x_i - \bar{x}| \cdot F_i}{n}$$

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada disciplina. O cálculo do desvio-médio será:

ESCORES			ALUNOS F_i	**xi**	**|xi-média|**	**|xi-média|*Fi**
35	|-	45	5	**40**	**22**	**111**
45	|-	55	12	**50**	**12**	**147**
55	|-	65	18	**60**	**2**	**40**
65	|-	75	14	**70**	**8**	**109**
75	|-	85	6	**80**	**18**	**107**
85	|-	95	3	**90**	**28**	**83**
TOTAL			58	**-**	**-**	**597**

Portanto,

$$\text{como } \bar{x} = 62{,}24$$

$$\text{então DM} = \frac{597}{58} = 10{,}29$$

Interpretação: Em média, a nota de cada aluno deste grupo teve um distanciamento de 10,29pontos em torno do desempenho médio deste grupo de alunos foi de 62,24 pontos nesta disciplina.

VARIÂNCIA E DESVIO-PADRÃO

A variância de um conjunto de dados é a média dos quadrados dos desvios dos valores a contar da média. A fórmula da variância poderá ser calculada de duas formas:

- POPULACIONAL, representada letra grega σ^2
- AMOSTRAL, representada por S^2.

1ª SITUAÇÃO: Dados não agrupados

Sejam os elementos $x_1, x_2, x_3, \ldots, x_n$, portanto "n" valores da variável X, com média igual a $\overline{x}$. A variância da variável aleatória de X é,

$$\sigma^2 = \frac{\sum (x_i - \mu)^2}{N} = \frac{1}{N} \cdot \left(\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{N} \right)$$

ou

$$S^2 = \frac{\sum (x_i - \overline{x})^2}{n - 1} = \frac{1}{n - 1} \cdot \left(\sum x_i^2 - \frac{\left(\sum x_i\right)^2}{n} \right)$$

Obs: A Segunda fórmula é chamada de "Fórmula Desenvolvida".

Exemplo:

Suponha o conjunto de tempo de serviço de cinco funcionários: 3, 7, 8, 10 e 11. Determinar o desvio-padrão deste conjunto de dados.

$$\text{como } \overline{x} = 7{,}8$$

$$\text{então } S^2 = \frac{(3 - 7{,}8)^2 + (7 - 7{,}8)^2 + (8 - 7{,}8)^2 + (10 - 7{,}8)^2 + (11 - 7{,}8)^2}{5 - 1} = \frac{38{,}8}{4} \Rightarrow S^2 = 9{,}7\text{anos}^2$$

Interpretação: encontramos então uma variância para o tempo de serviço de 9,7anos^2. Para eliminarmos o quadrado da unidade de medida, extraímos a raiz quadrada do resultado da variância, que chegamos a uma terceira medida de dispersão, chamada de **DESVIO-PADRÃO:**

- POPULACIONAL, representada letra grega $\sigma = \sqrt{\sigma^2}$
- AMOSTRAL, representada por $S = \sqrt{S^2}$

Portanto, o desvio-padrão do exemplo foi de 3,11anos. Ou seja, se calcularmos um intervalo utilizando um desvio-padrão em torno da média, encontraremos a concentração da maioria dos dados.

2ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por valores simples

Quando os dados estiverem agrupados numa distribuição de freqüência usaremos a variância dos valores x_1, $x_2, x_3, \ldots, x_n$, ponderados pelas respectivas freqüências absolutas: $F_1, F_2, F_3, \ldots, F_n$. Assim

$$\sigma^2 = \frac{\sum (x_i - \mu)^2 \cdot F_i}{N} = \frac{1}{N} \cdot \left(\sum x_i^2 \cdot F_i - \frac{\left(\sum x_i \cdot F_i\right)^2}{N} \right)$$

ou

$$S^2 = \frac{\sum (x_i - \bar{x})^2 \cdot F_i}{n - 1} = \frac{1}{n - 1} \cdot \left(\sum x_i^2 \cdot F_i - \frac{\left(\sum x_i \cdot F_i\right)^2}{n} \right)$$

Exemplo:

Em um determinado dia foi registrado o número de veículos negociados por uma amostra de 10 vendedores de uma agência de automóveis como mostra a tabela abaixo. O cálculo do desvio-médio será:

veículos negociados (x_i)	número de vendedores (F_i)	**(xi-média)^2**	**(xi-média)^2*Fi**
1	1	**2,56**	**2,56**
2	3	**0,36**	**1,08**
3	5	**0,16**	**0,80**
4	1	**1,96**	**1,96**
TOTAL	10	**5,04**	**6,40**

OU

veículos negociados (x_i)	número de vendedores (F_i)	**xi*Fi**	**xi^2*Fi**
1	1	**1**	**1**
2	3	**6**	**12**
3	5	**15**	**45**
4	1	**4**	**16**
TOTAL	10	**26**	**74**

$$\text{como } \bar{x} = 2,6$$
$$\text{então } S^2 = \frac{6,4}{9} = 0,71 \text{veículos}^2$$
$$\Rightarrow S = \sqrt{0,71 \text{veículos}^2} = 0,84 \text{veículos}$$

$$S^2 = \frac{1}{9} \cdot \left[74 - \frac{26^2}{10} \right] = 0,71 \text{veículos}^2$$
$$\Rightarrow S = \sqrt{0,71 \text{veículos}^2} = 0,84 \text{veículos}$$

<u>Interpretação</u>: Portanto, o desvio-padrão do exemplo foi de 0,84 veículos. Ou seja, se calcularmos um intervalo utilizando um desvio-padrão em torno da média, encontraremos a concentração da maioria dos veículos negociados por vendedor.

➪ 3ª SITUAÇÃO: Dados agrupados em uma distribuição de frequência por classes

Quando os dados estiverem agrupados numa distribuição de freqüência usaremos a variância dos pontos médios x_1, x_2, x_3,...,x_n de cada classe, ponderados pelas respectivas freqüências absolutas: F_1, F_2, F_3, ... , F_n. Desta forma, o cálculo da variância passa a ser igual ao da 2ª situação. Assim

$$\sigma^2 = \frac{\sum (x_i - \mu)^2 \cdot F_i}{N} = \frac{1}{N} \cdot \left(\sum x_i^2 \cdot F_i - \frac{\left(\sum x_i \cdot F_i\right)^2}{N} \right)$$

ou

$$S^2 = \frac{\sum (x_i - \bar{x})^2 \cdot F_i}{n - 1} = \frac{1}{n - 1} \cdot \left(\sum x_i^2 \cdot F_i - \frac{\left(\sum x_i \cdot F_i\right)^2}{n} \right)$$

Exemplo:

A tabela abaixo representa os escores obtidos por um grupo de 58 alunos matriculados em uma determinada disciplina. O cálculo do desvio-médio será:

ESCORES	ALUNOS F_i	xi	(xi-média)^2	(xi-média)^2*Fi
35 |- 45	5	**40**	**495**	**2.473**
45 |- 55	12	**50**	**150**	**1.798**
55 |- 65	18	**60**	**5**	**90**
65 |- 75	14	**70**	**60**	**843**
75 |- 85	6	**80**	**315**	**1.893**
85 |- 95	3	**90**	**771**	**2.312**
TOTAL	58	**-**	**-**	**9.409**

OU

ESCORES	ALUNOS F_i	xi	xi*Fi	xi^2*Fi
35 |- 45	5	**40**	**200**	**8.000**
45 |- 55	12	**50**	**600**	**30.000**
55 |- 65	18	**60**	**1.080**	**64.800**
65 |- 75	14	**70**	**980**	**68.600**
75 |- 85	6	**80**	**480**	**38.400**
85 |- 95	3	**90**	**270**	**24.300**
TOTAL	58	**-**	**3.610**	**234.100**

$$\text{como } \bar{x} = 62{,}24$$

$$\text{então } S^2 = \frac{9.409}{57} = 165{,}1\text{pontos}^2$$

$$\Rightarrow S = \sqrt{165{,}1\text{pontos}^2} = 12{,}85\text{pontos}$$

$$S^2 = \frac{1}{57}\cdot\left[234.100 - \frac{3.610^2}{58}\right] = 165{,}1\text{pontos}$$

$$\Rightarrow S = \sqrt{165{,}1\text{pontos}^2} = 12{,}85\text{pontos}$$

<u>Interpretação</u>: Portanto, o desvio-padrão do exemplo foi de 12,85 pontos. Ou seja, se calcularmos um intervalo utilizando um desvio-padrão em torno do escore médio de 62,24 pontos, encontraremos a concentração da maioria dos alunos dentro deste intervalo de pontuação.

COEFICIENTE DE VARIAÇÃO

Trata-se de uma média relativa à dispersão, útil para a comparação e observação em termos relativos do grau de concentração em torno da média de séries distintas. É dada por:

$$CV = \frac{\sigma}{\mu}\cdot 100 \quad \text{OU} \quad CV = \frac{S}{\bar{x}}\cdot 100$$

Classificação da distribuição quanto à dispersão:

- DISPERSÇÃO BAIXA: CV $\leq$ 15%
- DISPERSÇÃO MÉDIA: 15% $<$ CV $<$ 30%
- DISPERSÇÃO ALTA: CV $\geq$ 30%

Exemplo:

Numa empresa o salário médio dos funcionários do sexo masculino é de R$ 4.000,00, com um desvio padrão de R$ 1.500,00, e os funcionários do sexo feminino é em média de R$ 3.000,00, com um desvio padrão de R$ 1.200,00. Então:

$$\text{Sexo masculino: } CV = \frac{1500}{4000} \cdot 100 = 37{,}5\%$$

$$\text{Sexo feminino: } CV = \frac{1200}{3000} \cdot 100 = 40\%$$

Interpretação: Logo, podemos concluir que o salário das mulheres apresenta maior dispersão relativa que a dos homens. Para obtermos o resultado de C.V basta multiplicarmos por 100.

EXERCÍCIOS

1. Desvio Médio para o conjunto de dados abaixo será:

xi	Fi
5	2
7	3
8	5
9	4
11	2

a) ( ) 1,28 c) ( ) 1,00
b) ( ) 1,20 d) ( ) 0,83

2. O Desvio Padrão de um conjunto de dados é 9. A variância é:
a) ( ) 3 c) ( ) 81
b) ( ) 36 d) ( ) 18

3. Na distribuição de valores iguais, o Desvio padrão é:
a) ( ) negativo c) ( ) zero
b) ( ) a unidade d) ( ) positivo

4. O calculo da variância supõe o conhecimento da:
a) ( ) Fac c) ( ) mediana
b) ( ) média d) ( ) moda

5. A variância do conjunto de dados tabelados abaixo será:

Classes	Fi
03 |- 08	5
08 |- 13	15
13 |- 18	20
18 |- 23	10

a) ( ) 1,36 c) ( ) 4,54
b) ( ) 18,35 d) ( ) 20,66

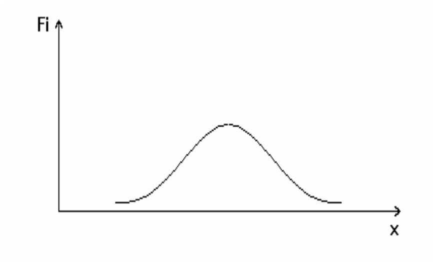
UNIDADE VII - MEDIDAS DE ASSIMETRIA E DE CURTOSE

MEDIDAS DE ASSIMETRIA

DEFINIÇÃO: grau de deformação de uma distribuição em relação ao eixo de simetria.

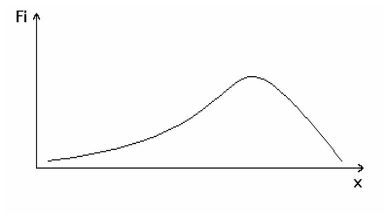
Podemos observar os tipos de assimetria abaixo:

a)



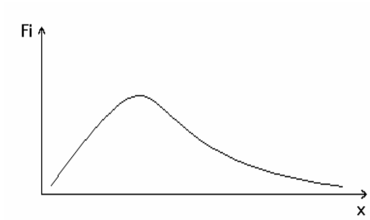
$\Rightarrow \bar{x} = Md = Mo$

b)



$\Rightarrow \bar{x} < Md < Mo$

c)



$\Rightarrow Mo < Md < \bar{x}$

Existem várias coeficientes com o objetivo de quantificar tais assimetrias. Estudaremos dois destes coeficientes que veremos a seguir:

➯ COEFICIENTE DE PEARSON

O coeficiente de Pearson é apresentado pela seguinte fórmula:

$$As = \frac{\mu - Mo}{\sigma} \quad \text{ou} \quad As = \frac{\bar{x} - Mo}{S}$$

Classificação do coeficiente de Pearson:

$As = 0$	DISTRIBUIÇÃO SIMÉTRICA
$0 < As < 1$	DISTRIBUIÇÃO ASSIMÉTRICA POSITIVA FRACA
$As \geq 1$	DISTRIBUIÇÃO ASSIMÉTRICA POSITIVA FORTE
$-1 < As < 0$	DISTRIBUIÇÃO ASSIMÉTRICA NEGATIVA FRACA
$As \leq -1$	DISTRIBUIÇÃO ASSIMÉTRICA NEGATIVA FORTE

↳ COEFICENTE DE BOWLEY

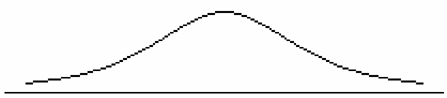
$$As = \frac{Q_3 + Q_1 - 2 \cdot Md}{Q_3 - Q_1}$$

Classificação do coeficiente de Bowley:

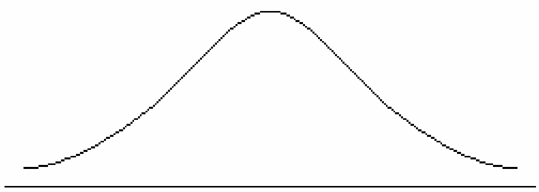
$As = 0$	DISTRIBUIÇÃO SIMÉTRICA
$0 < As \leq 0,1$	DISTRIBUIÇÃO ASSIMÉTRICA POSITIVA FRACA
$0,1 < As < 0,3$	DISTRIBUIÇÃO ASSIMÉTRICA POSITIVA MODERADA
$0,3 \leq As \leq 1$	DISTRIBUIÇÃO ASSIMÉTRICA POSITIVA FORTE
$-0,1 \leq As < 0$	DISTRIBUIÇÃO ASSIMÉTRICA NEGATIVA FRACA
$-0,3 < As < -0,1$	DISTRIBUIÇÃO ASSIMÉTRICA NEGATIVA MODERADA
$-1 \leq As \leq -0,3$	DISTRIBUIÇÃO ASSIMÉTRICA NEGATIVA FORTE

MEDIDA DE CURTOSE

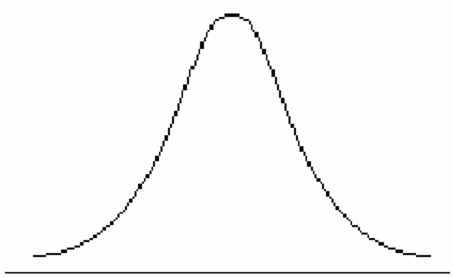
Entende-se por curtose o grau de achatamento de uma distribuição. Podemos ter:



⇒ CURVA PLATICÚRTICA



⇒ CURVA MESOCÚRTICA



⇒ CURVA LEPTOCÚRTICA

Para medir o grau de curtose utilizaremos o coeficiente

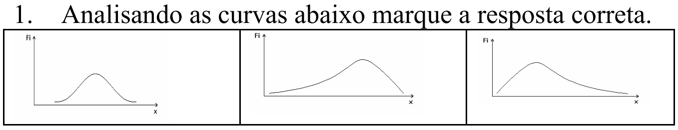
$$K = \frac{Q_3 - Q_1}{2 \cdot (P_{90} - P_{10})}$$

Classificação do coeficiente de Curtose:

$K = 0{,}263$	CURVA MESOCÚRTICA
$K > 0{,}263$	CURVA PLATICÚRTICA
$K < 0{,}263$	CURVA LEPTOCÚRTICA

EXERCÍCIOS

1. Analisando as curvas abaixo marque a resposta correta.



(I) (II) (III)

a) a curva I é simétrica - $\bar{x} > med > mo$;
b) a curva II é assimétrica positiva - $mo > \sigma^2 > \bar{x}$;
c) a curva I é simétrica $\bar{x} = med = mo$;
d) a curva III é simétrica positiva $\bar{x} = med = mo$;

2. Para as distribuições abaixo foram calculados

Distrib. A

Classes	Fi
02 |- 06	6
06 |- 10	12
10 |- 14	24
14 |- 18	12
18 |- 22	6

$\bar{x} = 12Kg$
$Med = 12Kg$
$Mo = 12Kg$
$S = 4{,}42Kg$

Distrib. B

Classes	Fi
02 |- 06	6
06 |- 10	12
10 |- 14	24
14 |- 18	30
18 |- 22	6

$\bar{x} = 12{,}9Kg$
$Med = 13{,}5Kg$
$Mo = 16Kg$
$S = 4{,}20Kg$

Distrib. C

Classes	Fi
02 |- 06	6
06 |- 10	30
10 |- 14	24
14 |- 18	12
18 |- 22	6

$\bar{x} = 11{,}1Kg$
$Med = 10{,}5Kg$
$Mo = 8Kg$
$S = 4{,}20Kg$

Marque a alternativa correta:
a) a distribuição I é assimétrica negativa;
b) a distribuição II é assimétrica positiva;
c) a distribuição III é assimétrica negativa moderada.
d) a distribuição I é simétrica;

3. Sabe-se que uma distribuição apresentou as seguintes medidas:

$Q_1 = 24{,}4cm$ $Q_3 = 41{,}2cm$
$P_{10}=20{,}2cm$ $P_{90} = 49.5cm$,

Com tais medidas a curtose é:
a) ( ) Leptocúrtica c) ( ) Mesocúrtica
b) ( ) Platicúrtica d) ( ) Assimétrica.

UNIDADE VIII – INTRODUÇÃO À TEORIA DA PROBABILIDADE

EXPERIMENTO ALEATÓRIO OU NÃO DETERMINISTICO - E

Definição:

1. É o processo de observação ou medida de um determinado fenômeno em estudo.
2. É o experimento que repetido sob as mesmas condições, conduz a resultados, em geral, distintos.

Exemplos:

E1 – lançamento de um dado e observar o número na face superior.

E2 – lançamento de uma moeda e observar o valor na face superior.

E3 – lançamento de um dado e uma moeda, nesta seqüência, observar os valores nas faces superiores.

E4 – um casal deseja ter três filhos e observar o sexo, de acordo com a ordem de nascimentos das crianças.

ESPAÇO AMOSTRAL - S

Definição:

Um espaço amostral é um conjunto de todas as ocorrências possíveis de um determinado experimento aleatório E.

Exemplos: Considere os experimentos aleatórios apresentados anteriormente:

No E1 - S={1, 2, 3, 4, 5, 6}

No E2 - S={k, c}, onde k=cara, C=coroa.

No E3 - S={1k, 2k, 3k, 4k, 5k, 6k, 1c, 2c, 3c, 4c, 5c, 6c}

No E4 - S={MMM, MMF, MFM, MFF, FMM, FMF, FFM, FFF}

EVENTOS – (qualquer letra maiúscula do alfabeto)

Definição:

Um evento é qualquer subconjunto de ocorrências de um determinado espaço amostral S.

Exemplo: Considere o experimento aleatório E3, com seu respectivo espaço amostral S:

S={1k, 2k, 3k, 4k, 5k, 6k, 1c, 2c, 3c, 4c, 5c, 6c}

Determine os seguintes eventos:

A = ocorrência de valor cara (K)

B = ocorrência de valor par

C = ocorrência de valor coroa (C)

D = ocorrência de valor ímpar

E = ocorrência de número primo

F = ocorrência de valor maior que 4

G = ocorrência de valor menor ou igual a 3

H = ocorrência de valor par **ou** cara (K)

I = ocorrência de valor par **ou** ímpar

J = ocorrência de valor par **e** cara (K)

K = ocorrência de valor par **e** ímpar

L = ocorrência de valor maior que 7

TIPOS DE EVENTOS

- **EVENTO CERTO**

Definição:

É aquele evento que se igual ao espaço amostral S.

Exemplo: O **evento I** acima é um evento certo.

EVENTO IMPOSSÍVEL

Definição:

É aquele evento que não possui elemento algum.

Exemplo: Os **eventos K e L** acima são eventos impossíveis.

EVENTOS MUTUAMENTES EXCLUSIVOS

Definição:

Dois eventos A e B quaisquer são chamados de mutuamente exclusivos, se eles não podem ocorrer simultaneamente, isto é,

$$\mathbf{A \cap B = \emptyset}$$

Exemplo: Considere os eventos descritos acima:

Os eventos **A** e **C** são mutuamente exclusivos, pois $A \cap C = \emptyset$.

Os eventos **B** e **D** são mutuamente exclusivos, pois $B \cap D = \emptyset$.

Os eventos **C** e **J** são mutuamente exclusivos, pois $C \cap J = \emptyset$.

Os eventos **H** e **J** não são mutuamente exclusivos, pois $H \cap J \neq \emptyset$.

EVENTOS COMPLEMENTARES

Definição:

Dois eventos A e B quaisquer são chamados de complementares se:

$$\mathbf{A \cap B = \emptyset}$$

$$\mathbf{A \cup B = S}$$

Exemplo: Considere os eventos descritos no exemplo acima:

Os eventos **A** e **C** são complementares, pois $A \cap C = \emptyset$ e $A \cup C = S$.

Os eventos **B** e **D** são complementares, pois $B \cap D = \emptyset$ e $B \cup D = S$.

Os eventos **H** e **J** não são complementares, pois $H \cap J \neq \emptyset$ e $H \cup J \neq S$.

Os eventos **F** e **K** não são complementares, pois $F \cap K \neq \emptyset$ apesar de $F \cup K = S$.

Os eventos **C** e **J** não são complementares, pois $C \cup J \neq S$ apesar de $C \cap J = \emptyset$.

PROBABILIDADE:

Enfoque Teórico

A probabilidade de ocorrência de um evento A, P(A), é um número real que satisfaz as seguintes condições:

a) $0 \leq P(A) \leq 1$

b) $P(S) = 1$

c) Se A e B são **eventos mutuamente exclusivos** então $P(A \cup B) = P(A) + P(B)$

d) Se $A_1, A_2, ..., A$, ... São mutuamente exclusivos, dois a dois, então:

Principais teoremas:

I) $P(\overline{A}) = 1 - P(A)$

II) Se A é um evento impossível de ocorrer ($A=\emptyset$), então $P(A) = P(\emptyset) = 0$.

III) Se A e B são eventos quaisquer, então: $P(A \cup B) = P(A) + P(B) - P(B \cap A)$.

CÁLCULO DA PROBABILIDADE

A probabilidade deverá ser calculada a partir da fórmula: $P(A) = \dfrac{n(A)}{n(S)}$

Exemplo:

Seja o Experimento E o lançamento de um dado e o seu espaço amostral dado por: S = {1, 2, 3, 4, 5, 6}. Qual a probabilidade do evento A – Números maiores e iguais a 2?

O Evento A pode ser descrito na forma: A ={2, 3, 4, 5, 6}

n(A) = 5 e n(S) = 6. Logo a probabilidade do evento A é $P(A) = \dfrac{n(A)}{n(S)} = 5/6$.

PROBABILIDADE CONDICIONAL

Ilustração:

Seja o experimento aleatório E: lançar um dado e o evento A = {sair o número 3}. Então:

$$P(A) = \frac{1}{6}$$

Seja o evento B = {sair o número impar} = {1, 3, 5}

Podemos estar interessados em avaliar a probabilidade do evento A estar condicionado à ocorrência do evento B, designado por P(A|B), onde o evento A é o evento condicionado e o evento B o condicionante.

Assim $P(A|B) = \dfrac{1}{3}$

Formalmente a probabilidade condicionada é definida por:

"Dado dois eventos quaisquer A e B, denotaremos P(A|B), por".

$$P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{n(A\cap B)}{n(B)},$$

Com $P(B)\neq 0$, pois B já ocorreu.

TEOREMA DO PRODUTO

A probabilidade da ocorrência simultânea de dois eventos quaisquer A e B, do mesmo espaço amostra, é igual ao produto da probabilidade de ocorrência do primeiro deles pela probabilidade condicional do outro, dado que o primeiro ocorreu.

Assim:

$$P(A|B)=\frac{P(A\cap B)}{P(B)} \quad \Rightarrow \quad P(A\cap B)=P(B)\cdot P(A|B)$$

INDEPENDÊNCIA ESTATÍSTICA

Um evento A é considerado independente de um outro evento B se a probabilidade de A é igual à probabilidade condicional de A dado B, isto é, se:

$$P(A)=P(A|B)$$

Considerando o teorema do produto podemos afirmar que:

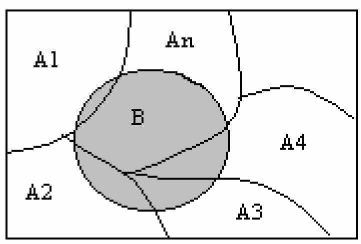
$$P(A\cap B)=P(A)\cdot P(B)$$

TEOREMA DE BAYES

Suponha que os eventos A_1, A_2, ...,A_n formam uma partição de um espaço amostral S; ou seja, os eventos A_i são mutuamente exclusivos e sua união é S. Seja B outro evento qualquer. Então:

$$B = S\cap B = (A_1\cup A_2\cup...\cup A_n)\cap B = (A_1\cap B)\cup(A_2\cap B)\cup...\cup(A_n\cap B)$$

Onde os $A_i\cap B$ são também mutuamente exclusivos.



Consequentemente,

$$P(B) = P(A_1\cap B) + P(A_2\cap B) +... +P(A_n\cap B)$$

Assim pelo teorema da multiplicação,

$$P(B) = P(A_1)P(B\backslash A_1)+ P(A_2)P(B\backslash A_2)+...+ P(A_N)P(B\backslash A_N)$$

Por outro lado, para qualquer i, a probabilidade condicional de A_i dado B é definida como

$$P(A_i \backslash B) = \frac{P(A_i \cap B)}{P(B)}$$

Nesta equação, usamos (1) para substituir P(B) e $P(A_i \cap B) = P(A_i)P(B\backslash A_i)$ para substituir $P(A_i \cap B)$, obtendo assim o:

Teorema de Bayes: Suponha $A_1, A_2, ..., A_n$ ser uma partição de S e B, um evento qualquer. Então, para qualquer i,

$$P(A_i \backslash B) = \frac{P(A_i)P(B \backslash A_i)}{P(A_1)P(B \backslash A_1) + P(A_2)P(B \backslash A_2) + ... + P(A_n)P(B \backslash A_n)}$$

Exemplos:

Três máquinas, A, B e C produzem 50%, 30% e 20%, respectivamente do total de peças de uma fábrica. As percentagens de produção defeituosa destas máquinas são 3%, 4% e 5%. Se uma peça é selecionada aleatoriamente, ache a probabilidade de ela ser defeituosa. Suponha agora que uma peça selecionada aleatoriamente seja defeituosa. Encontre a probabilidade de ela ter sido produzida pela máquina A

EXERCÍCIOS

1. Lance um dado e uma moeda um após o outro nesta seqüência.

a) Construa o espaço amostral
b) Enumere os resultados seguintes
   I. A ={coroa marcada por par}
   II. B = {cara marcada por ímpar}
   III. C = {Múltiplo de 3}
c) Expresse os eventos
   I. B complementar
   II. A ou B ocorrem
   III. B ou C ocorrem
   IV. A ou B complementar
c) Calcule as probabilidades abaixo:

P(A), P(B), P(C), P($\overline{A}$),P($\overline{B}$), P($A \cup B$) e P($B \cup C$)

2. Um revendedor de carros tem dois carros, corsas 1996, na sua loja para serem vendidos, interessa-nos saber quanto cada um dos dois vendedores venderá ao final de uma semana. Como representar "o primeiro vendedor não vende nenhum carro" e depois "o segundo vendedor vende ao menos um dos carros".

3. Se A é o evento "Um estudante fica em casa para estudar". E B é o evento "o estudante vai ao cinema", P(A) = 0,64 e P(B) = 0,21. Determine:

$P(A^c)$, $P(B^c)$, P(B/A)

4. Se P(A) = ½ ; P(B) = ¼ e A e B são mutuamente exclusivos então:
a) $P(A^c)$
b) $P(B^c)$
c) $P(A \cap B)$

5. Se P(A) = ½; P(B) = 1/3 e $P(A \cap B)$ = ¼ calcule P(AUB).

6. Quantas comissões de três pessoas podem formar com um grupo de 10 pessoas?

7. A probabilidade de três jogadores acertarem um pênalti são respectivamente 2/3, 4/5 e 7/10. Se cada um cobrar uma única vez, qual a probabilidade de:

a) Todos acertarem
b) Ao menos um acertar
c) Nenhum acertar

8. Qual a probabilidade de duas pessoas aniversariarem no mesmo dia da semana?

9. Sr Ray Moon Dee, ao dirigir-se ao trabalho, usa um ônibus ou o metrô com probabilidade de 0,2 e 0,8, nessa ordem. Quando toma o ônibus, chega atrasado 30% das vezes. Quando toma o metrô, atrasa-se 20% dos dias. Se o Sr Ray Moon Dee Chegar atrasado ao trabalho em determinado dia, qual a probabilidade dele haver tomado um ônibus?

10. Em certo colégio, 5% dos homens e 2% das mulheres tem mais que 1,80m de altura. Por outro lado, 60% dos estudantes são homens. Se um estudante é selecionado aleatoriamente e tem mais de 1,80m de altura, qual a probabilidade de que o estudante seja mulher?