

1. ORIGEM E HISTÓRIA DA ESTATÍSTICA

A origem da palavra Estatística do ponto de vista etimológico vem do vocábulo latino “Status”. Originalmente essa palavra significava informações de interesse ou úteis ao “Estado”, ou País, seja no que se refere à cobrança de contribuições, taxas e impostos, seja também para o conhecimento dos recursos naturais de um país, da composição da população, etc.(CONAGIN, 1961).

Historicamente, operações de contagem de pessoas, animais e avaliação de bens pessoais, isto é, atividades que hoje fazem parte de um censo populacional e econômico, eram já realizadas, de modo incipiente, no Egito do tempo dos faraós, na Grécia antiga e no império Romano como atividades de interesse do estado, seja com enfoque econômico associado à cobrança de impostos, seja para o controle da produção e distribuição de alimentos e/ou outros tipos de riquezas, ou por questões de segurança, no sentido de detectar quantos eram e onde estavam os potenciais inimigos representados pelos povos subjugados. Enfim, por alguma razão, sempre houve interesse em se conhecer o status populacional, daí a origem da palavra Estatística.

Segundo Yule e Kendall (1958), as expressões *Statistic*, *Statist* e *Statistical* parecem ter sido derivadas do latim *Status*, com duplo significado i) Estado político e ii) Situação das coisas em determinado momento – *Statu quo*.

O nascimento da palavra Estatística também pode ser atribuído como sendo proveniente do alfabeto latim denominado *statisticum*, que significa atividades relativas ao Estado (DROESBEKE E TASSI, 1990; DUTARTE E PIEDNOIR, 2001; JOZEAU, 2001). Porém não existe na literatura um consenso sobre quem ou qual civilização utilizou primeiramente o termo estatística, tal como usamos nos dias atuais, ou seja, para designar uma área do conhecimento ou ainda um conjunto de medidas-resumo de um conjunto de dados, ou também um conjunto de técnicas utilizadas para realizar inferência estatística ou conclusões sobre uma população.

A expressão estatística vem do idioma alemão *statistik* sendo adotada pelo acadêmico alemão Gottfried Achenwall¹ (1719-1772) na cidade de Göttingen na Alemanha, em 1749, que foi um notável continuador dos estudos de Hermann Conring² (1606-1681).

O verbete “*statistic*” é descrito na Enciclopédia Britânica em 1797.

Em 3000 a.C., embora nem existisse a palavra escrita como a conhecemos hoje, há indícios que já se faziam censo na Babilônia, Egito e na China. No antigo testamento o livro dos números mostra a instrução a Moisés para fazer um levantamento dos homens de Israel aptos para guerrear.

Desde a remota Antiguidade, os governos têm se interessado por informações sobre suas populações e riquezas, tendo em vista, principalmente, fins militares e tributários. O

¹ Gottfried Achenwall (1719-1772) economista, professor na universidade de Göttingen, considerado pelos alemães como o pai da estatística (nasceu na cidade de [Elbing](#), [20 de outubro](#) de [1719](#), e faleceu na cidade de [göttingen](#) em [1º de maio](#) de [1772](#) na Alemanha, foi um [historiador](#) e [jurista alemão](#), é considerado um dos intelectuais que mais significativamente contribuiu para o desenvolvimento da estatística, a qual é uma ciência que guarda muito poucas relações com a atual [estatística](#), pois concernia à descrição abrangente das características sócio-político-econômicas dos diferentes estados. Achenwall estudou nas universidades de [Jena](#), [Halle](#) e [Leipzig](#). Em [1748](#), foi nomeado docente pela [universidade de Göttingen](#); em [1761](#), obteria nesta mesma universidade o cargo de professor de [direito natural](#) e [política](#).

² Hermann Conring ([9 de novembro](#) de [1606](#) - [12 de dezembro](#) de [1681](#)) foi um intelectual alemão nascido no norte da Alemanha, o qual foi responsável por significativas contribuições no estudo da medicina, na política e da lei. Conring mostrou cedo promessa como um aluno. Durante a sua vida como professor no norte da Alemanha, Conring abordou primeiro a medicina, produzindo importantes estudos sobre a circulação sanguínea, e mais tarde em sua carreira ele abordou a política. Conring nasceu em [Norden](#), uma cidade costeira no [oriente frísias](#), um território como muitas outras áreas que mais tarde iria se tornar a Alemanha.

registro de informações perde-se no tempo. O filósofo e pensador chinês Confúcio³ relatou levantamentos feitos na China, há mais de 2000 anos antes da era cristã no livro sagrado “*CHOUKING*” no qual se fizeram constar os dados estatísticos do grande império chinês da sua época. No antigo Egito, os faraós fizeram uso sistemático de informações de caráter estatístico, conforme evidenciaram pesquisas arqueológicas. Desses registros também se utilizaram as civilizações pré colombianas dos maias, astecas e incas. É conhecido de todos os cristãos o recenseamento dos judeus, ordenado pelo Imperador Augusto. Os balancetes do império romano, o inventário das posses de Carlos Magno, o *Doomsday Book*, registro que Guilherme, o Conquistador, invasor normando da Inglaterra, no século 11, mandou levantar das propriedades rurais dos conquistados anglo-saxões para se inteirar de suas riquezas, são alguns exemplos anteriores à emergência da estatística descritiva no século 16, na Itália. Essa prática tem sido continuada nos tempos modernos, por meio dos recenseamentos, dos quais temos um exemplo naquele que se efetua a cada decênio, em nosso País, pela Fundação IBGE, órgão responsável por nossas estatísticas (dados estatísticos) oficiais.

Desde as administrações antigas, a estatística sempre se faz presente a favor do estado, a ponto de herdar o nome do vocábulo “*Status*”, que quer dizer “Estado”, servindo de apoio, pelo seu aspecto quantitativo, para uma administração consciente, quer fosse em época de paz quer seja em período de guerra, em cujos campos ela prestava o seu trabalho, em consonância com a imagem fria, mas verdadeira dos números que sempre lhe coube analisar, no campo dos fenômenos coletivos, econômicos sociais ou científicos, geralmente com a finalidade de controle fiscal ou de segurança nacional. Sendo assim, percebe-se que na antiguidade em várias civilizações já se registravam o número de habitantes, nascimentos, óbitos, faziam estimativas pertinentes às riquezas individual e social, cobravam impostos, taxas etc., podendo todas essas ações serem chamadas de estatísticas.

Entretanto, a origem da estatística científica teve início na metade do século 17, quando passou a ser considerada como uma disciplina autônoma tendo como objetivo a descrição das coisas do Estado e como fundador Herman Conring (1606 - 1681). Seus seguidores foram, dentre outros, Martin Schymeitzel⁴ (1679 - 1747) e Gottfried Achenwall (1719 - 1782), chegando mais tarde este último a superar a fama do próprio Conring. Foi graças a Gottfried Achenwall que a estatística firmou-se como ciência determinando o seu objetivo e suas relações com as demais (SILVEIRA JÚNIOR, et al. 1989).

Um marco significativo no desenvolvimento da estatística foram as publicações *Natural and Political Annotations Made Upon the Bills of Mortality* (1666) e *Five Essays on Political Arithmetick* (1683) publicados, respectivamente, por John Graunt (1620-1674) e Willyam Petty (1623-1687). Graunt e Petty, cidadãos ingleses, não se limitaram a simples apresentação de totais e estabeleceram relações entre esses números dando início ao que hoje se chama demografia (SILVEIRA JÚNIOR, et al. 1989).

O avanço no estudo do cálculo de probabilidades permitiu que a estatística fosse organicamente estruturada e fez com que se ampliasse largamente o seu campo de ação. Nessa mesma época iniciou-se o cálculo de probabilidades derivado do interesse do matemático e filósofo francês Blaise Pascal (1623-1662) em resolver jogos de azar. Data de 1713, entretanto, o primeiro grande tratado no campo da probabilidade, escrito pelo suíço Jacques Bernoulli (1654-1705), denominado *Ars Conjectandi* (Arte das conjecturas). Bernoulli exemplificou seu trabalho principalmente em termos de jogos de azar e na mesma linha seguiu alguns trabalhos subseqüentes, com destaque para a obra de Abraham de Moivre (1667-1754) iniciada no princípio do século XVIII, em 1718, com *Doctrine of Chances*. A famosa distribuição normal foi

³ Confúcio viveu entre os anos de [551 a.c.](#) a [479 a.c.](#), nome [latino](#) do pensador chinês Kung-Fu-Tze. foi a figura histórica mais conhecida na [china](#) como mestre, [filósofo](#) e teórico [político](#). Sua doutrina, o [confucionismo](#), teve forte influência não apenas sobre a [china](#) mas também sobre toda a [ásia](#) oriental. Conhece-se muito pouco da sua vida. Parece que os seus antepassados foram gente [nobre](#), mas o filósofo e [moralista](#) viveu pobre, e desde a infância teve de ser mestre de si mesmo. Na sua época, a china estava praticamente dividida em reinos [feudais](#) cujos senhores dependiam muito pouco do rei.

⁴ Intelectual alemão, professor de história e direito.

deduzida pelo francês de Moivre como resultado do limite da expansão do Binômio $(a + b)^n$, embora ainda não tenha sido colocada realmente como uma distribuição de probabilidade. Mas tarde, essa relação veio a ter uma importância muito grande, pois estabelece a aproximação da distribuição binomial por meio da distribuição normal (SILVEIRA JÚNIOR, et al. 1989).

O início do extraordinário avanço da teoria das probabilidades viria por meio do matemático francês Pierre Simon Laplace (1749-1827) com o seu clássico trabalho *Theorie analytique des probabilités* (1812) resultado de uma série de artigos sobre esse assunto. Nesse

trabalho ele mostra que a área sob a curva normal é $\sqrt{\pi}$, ou $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$. Ainda na sua

obra *Theorie analytique*, Laplace fornece uma prova formal do método dos quadrados mínimos já publicado pelo matemático francês Adrien-Marie Legendre (1752 - 1833) (SILVEIRA JÚNIOR, et al. 1989).

Praticamente no mesmo período, o matemático alemão Johann Karl Friedrich Gauss (1777-1855) chegou aos mesmos resultados sobre a curva normal de probabilidades, estudando a distribuição dos erros de medida.

Deve-se ao astrônomo belga Lambert Adolphe Jacques Quételet (1796-1874), principalmente, a relação entre a Estatística e o cálculo de probabilidades. Quételet por meio de estudos na área social mostrou que muitos fenômenos vivos possuíam um comportamento regular. A expressão matemática dessa regularidade é o que hoje se conhece como distribuição de probabilidade (SILVEIRA JÚNIOR, et al. 1989).

Após Quételet, a Estatística teve um desenvolvimento sem precedentes, sendo o fenômeno da regularidade observado em muitos campos da pesquisa. As distribuições de probabilidade começaram a ser deduzidas e com isto aumentava ainda mais o campo de aplicação. O estatístico Karl Pearson (1857-1936), em 1890 iniciou o estudo sobre o relacionamento entre variáveis, estimulado pelos trabalhos do estatístico Inglês Francis Galton (1822-1911). Galton, 1900 deduziu a distribuição de Qui-Quadrado, dentro de um contexto de ajustamento à distribuições, embora o matemático alemão Friedrich Robert Helmert (1884-1917) já a houvesse deduzido em 1875, trabalhando com distribuições amostrais de variância (SILVEIRA JÚNIOR, et al. 1989).

O cidadão inglês William Sealey Gosset (1876-1937), aluno de Karl Pearson e conhecido também pelo pseudônimo de Student, descobriu em 1908 a distribuição "t" no intuito de resolver problemas relativos a pequenas amostras ($n < 30$ elementos).

Em 1890, nascia na Inglaterra Ronald Aymer Fisher (1890-1962) que haveria de contribuir de forma extremamente valiosa com a estatística. Fisher, com os resultados de Gosset, descobriu rapidamente as distribuições amostrais dos coeficientes de correlação, da regressão, da correlação múltipla e a distribuição da razão entre duas variâncias (SILVEIRA JÚNIOR, et al. 1989).

O período de 1920 a 1940 foi ainda mais extremamente fecundo. Fisher preocupou-se com o fato de que em situações experimentais, uma variável era explicada por várias outras, correlacionadas entre si, o que tornava impossível o estudo isolado de cada uma delas. Para resolver esse problema ele desenvolveu na estação experimental de Rothamstead, localizada próxima a Londres, esquemas experimentais de modo que os efeitos pudessem ser estudados independentemente e, no mesmo período, ele generalizou e deu ideia mais precisa à técnica chamada de análise de variância, até hoje uma das mais poderosas técnicas utilizada na estatística. Dois campos importantíssimos emergiram ainda no período posterior a 1925, o primeiro deles foi a teoria dos testes de hipóteses sob inspiração do inglês Egon Pearson (1895 - 1980) e Jerzy Neyman (1894 - 1981), e a teoria da estimação desenvolvida por Ronald Aymer Fisher. Fisher, antes de falecer na Austrália em 1962, contribuiu ainda com valiosos trabalhos a outras áreas do conhecimento, em especial à genética e, dentro desta, à genética humana (SILVEIRA JÚNIOR, et al. 1989).

Modernamente a estatística passou a desempenhar a função de auxiliar o método científico, isto é, vem prestando sua colaboração no setor do planejamento experimental, na elaboração de técnicas para a coleta de dados e participando na interpretação analítica dos

experimentos (análise de dados experimentais) e na avaliação dos parâmetros dos modelos utilizados na pesquisa em ciências físicas, biológicas e sociais, na tecnologia industrial, etc. (CONAGIN, 1961).

Na década de 40, a estatística teórica podia ser compreendida por alguém com conhecimento razoável em Matemática. A situação mudou muito, a ponto de a maioria dos estatísticos não conseguir, atualmente, ler os artigos publicados, dado seu alto grau de sofisticação em matemática, mesmo nas revistas supostamente de caráter aplicado. Entretanto, deve ser dito que a Estatística não é propriamente Matemática, nem mesmo matemática aplicada. Como lida com a coleta, a análise e a interpretação de dados, inclui, naturalmente, muita conjectura sagaz, diferente do rigor da demonstração matemática, para não mencionar o raciocínio indutivo envolvido na inferência estatística. Evidentemente, saber Matemática é importante para um estatístico e quanto mais, melhor, pois a teoria estatística não envolve apenas conceitos, necessitando também ser formalizada. Contudo, conhecer Matemática, embora necessário, não é suficiente para formar um estatístico. Na era atual a estatística caracteriza-se pelo aumento gradativo de matematização da estatística e da influência crescente do uso da computação eletrônica, ferramenta valiosíssima que permitiu que a estatística alargasse ainda mais os seus horizontes e campos de aplicação (SILVEIRA JÚNIOR, et al. 1989).

2 DEFINIÇÕES DE ESTATÍSTICA

2.1 DEFINIÇÃO

É a ciência que está interessada nos métodos científicos para a coleta, organização, resumo, apresentação e análise dos dados, bem como na obtenção de conclusões válidas e na tomada de decisões baseadas em tais análises. (Memória, 1960).

2.2 DEFINIÇÃO

É a matemática aplicada aos dados de observação, definição esta dada por Sir Ronald Aylmer Fisher ⁵(1890-1962).

2.3 DEFINIÇÃO

É um conjunto de métodos especialmente apropriado ao tratamento de dados numéricos afetados por uma multiplicidade de causas sendo esta definição devido a George Udny Yule ⁶(1871 - 1951) e Maurice George Kendall ⁷(1907 - 1983).

⁵ Sir Ronald Aylmer Fisher nasceu em [17 de fevereiro](#) de [1890](#) e morreu em [29 de julho](#) de [1962](#), foi um [estatístico](#), [biólogo evolucionário](#) e [geneticista inglês](#). foi descrito por [Anders Hald](#) como "um gênio que criou praticamente sozinho as fundações para a moderna ciência estatística, [Richard Dawkins](#) o descreveu como o maior dos sucessores de [Darwin](#), complementando a medida de informação própria para medir incerteza sobre espaços desordenados, proposta com sucesso pela [teoria da informação](#) e publicada em [1949](#) por [Claude Shannon \(1916-2001\)](#) e [Warren Weaver \(1894-1978\)](#), no livro [teoria matemática da comunicação](#) (The mathematical theory of communication), Fisher criou uma medida alternativa de informação apropriada para medir incerteza sobre espaços ordenados.

⁶ Nascido em 18 de fevereiro de 1871 em Morham na Escócia e morreu em 26 de junho de 1951, em Cambridge, Cambridgeshire na Inglaterra depois de estudar engenharia e física, George Yule trabalhou com Pearson em 1893. Alguns de seus primeiros trabalhos aparecem como exemplos nos artigos de Pearson. Yule trabalhou na University College London, onde foi promovido para professor assistente de matemática aplicada em 1896. Yule é o autor do trabalho intitulado sobre a teoria da correlação sido publicado pela primeira vez em 1897. Ele desenvolveu a sua abordagem junto com a teoria da regressão durante vários anos, conceituando novas aplicações do método dos mínimos quadrados e em 1920 a sua abordagem predominou nas aplicações em ciências sociais.

⁷ Estatístico inglês.

2.4 DEFINIÇÃO

É o estudo numérico dos fatos sociais, definição devida a Pierre Émile Levasseur⁸ (1828 – 1911).

2.5 DEFINIÇÃO

A estatística é a coleta, apresentação, análise e interpretação de dados numéricos, definição atribuída a Frederick Emory Croxton & Didley Johnstone Cowden⁹(1964).

3 IMPORTÂNCIA DA ESTATÍSTICA

A estatística é uma importante ferramenta que pode e deve ser empregada pelos pesquisadores em geral em diferentes situações, como pode ser visto, por exemplo, nos itens abaixo.

- i. Planejamento, análise e interpretação de experiências agropecuárias, biológicas, sociais e industriais;
- ii. Estudo de fenômenos hidrológicos e meteorológicos;
- iii. Previsão de população para efeito de projetos de interesse da comunidade rural e urbana;
- iv. Controle de qualidade de produtos agrícolas e industriais;
- v. Previsão de safras;
- vi. Estimativas de produção;
- vii. Estudo de fenômenos econômicos e sociais;
- viii. Estudos genéticos;
- ix. Simulação;
- x. Geoestatística;
- xi. Seguros;
- xii. ¹⁰Atuária;
- xiii. Medicina;
- xiv. Demografia;
- xv. Meio ambiente;
- xvi. Contabilidade e área fiscal;
- xvii. Física;
- xviii. Química (quimiometria);
- xix. Engenharia civil: construção de pontes, de viadutos etc.
- xx. Engenharia do petróleo e de energia;
- xxi. Engenharia sanitária;
- xxii. Engenharia florestal;
- xxiii. Ecologia e Arborização urbana;
- xxiv. etc.

4 DIVISÃO DA ESTATÍSTICA

Para estabelecer o âmbito dos estudos da disciplina e adotando-se um esquema prático de raciocínio, pode-se dizer que a estatística compreende duas funções ou campos bem amplos.

4.1 ESTATÍSTICA DESCRITIVA OU ESTATÍSTICA DEDUTIVA

É a parte da estatística que cuida da apresentação dos dados de observação por meio do uso de tabelas, de gráficos e de medidas resumo as quais são: as medidas de posição, de dispersão, de assimetria e de curtose.

⁸ Pierre Émile Levasseur (1828 – 1911), historiador, geógrafo, economista e estatístico francês, nasceu e faleceu em Paris.

⁹ Estatísticos americanos.

4.2 ESTATÍSTICA INDUTIVA, ESTATÍSTICA ANALÍTICA OU INFERÊNCIA ESTATÍSTICA

É a parte que tem por objetivo fornecer os métodos que propiciem a realização de inferências sobre uma população a partir de amostras delas provenientes, tendo por base o cálculo de probabilidades.

¹⁰Atuário é um especialista em avaliar riscos nos segmentos de Seguros, Capitalização, Previdência Social, Previdência Complementar, Instituições Financeiras, Bancos e Órgãos Oficiais de Resseguros e Previdência;

5 CONCEITOS BÁSICOS

5.1 POPULAÇÃO OU UNIVERSO ESTATÍSTICO (N)

No sentido geral é um conjunto de objetos, pessoas, acontecimentos, medidas etc. com pelo menos uma característica comum, cujo comportamento interessa analisar ou inferir. As características da população são comumente chamadas de parâmetros, os quais são valores fixos e ordinariamente desconhecidos, tendo como símbolo os caracteres gregos. Exemplo: As áreas em hectares das propriedades agrícolas produtoras de milho no Brasil.

5.2 TIPOS DE POPULAÇÃO

5.2.1 POPULAÇÃO FINITA

É aquela em que o número de elementos que a compõem não é muito grande, ou seja, é limitada em tamanho ou ainda quando $n/N \geq 0,05$. Exemplo: O número de frutos por árvore de mangueira contados em 1000 árvores de um pomar.

5.2.2 POPULAÇÃO INFINITA

É aquela em que o número de elementos que a compõem é muito grande, admitindo-se como infinito, ou ainda quando $n/N < 0,05$. Exemplo: As possíveis vazões medidas em m^3/s em uma seção do rio Mossoró no período de um ano.

5.3 AMOSTRA (n)

É um subconjunto de observações, indivíduos, objetos etc. selecionado de uma população, necessariamente finita, e que preserva as suas características básicas, cujo comportamento interessa analisar ou inferir. As características da amostra são chamadas de estatísticas, sendo simbolizadas por caracteres latinos. Exemplos: valores de volume de chuva em milímetros (mm) em Mossoró nos dez primeiros dias do mês de janeiro de 2007, valores de produção leiteira em quilogramas (Kg) do rebanho bovino holandês preto e branco, da Universidade Federal Rural do Semi Árido (UFERSA), durante o mês de Janeiro de 2008.

5.4 VARIÁVEL (X_j)

Característica populacional mensurável de acordo com alguma escala e que pode assumir diferentes valores de elemento para elemento. A variável será representada por letras maiúsculas. Tipo: X, Y, Z e W. Exemplo: A temperatura corporal de animais em graus Celsius;

5.5 CLASSIFICAÇÃO DAS VARIÁVEIS (TIPOS)

Os dados de observação ou dados estatísticos constituem a matéria prima para o trabalho estatístico, eles têm sua origem, seja por medição, por contagem ou pesagem, no estudo de atributo ou característica de plantas, animais ou qualquer objeto ou elemento de uma população ou amostra, como exemplo podem ser citados, a altura de plantas, o rendimento de uma cultura, o ganho de peso de bovinos, a renda anual de uma comunidade rural, as precipitações pluviométricas de uma região, as possíveis vazões de um rio etc.

Tais medidas são variáveis, porque originam valores que tendem a exibir certo grau de variabilidade quando se fazem mensurações sucessivas (tais valores são chamados domínio da

variável). Ao contrário da constante a qual é uma característica que deve assumir valores iguais de indivíduo para indivíduo.

Na maioria das vezes, a escolha do processo a se utilizar na análise ou descrição de dados estatísticos depende do tipo de dado considerado.

Existem 4 tipos de variáveis ou dados, inclusive oriundos de uma mesma população ou amostra, vale salientar que dados ou observações são os valores numéricos assumidos pela variável.

O tratamento dos dados depende dessa classificação, ou seja, às técnicas estatísticas utilizadas para dados provenientes de variáveis qualitativas são diferentes das técnicas utilizadas para dados provenientes de variáveis quantitativas.

Sendo assim as variáveis podem ser classificadas da seguinte maneira:

5.5.1 VARIÁVEIS QUALITATIVAS

São aquelas em que as modalidades ou categorias formam um conjunto sem forma definida, não estruturado numericamente, ou seja, quando não há ligação entre esses atributos, independentemente do fato de constituírem um conjunto completo de observações de um estudo. Ou ainda são aquelas que apresentam como possíveis realizações uma qualidade ou atributo do elemento pesquisado, por isso não são inerentemente numéricas e devem ser convertidas em valores numéricos antes do processamento estatístico. Exemplo: a cor da pelagem de animais de um rebanho caprino, enquanto que, branco, preto, pardo, malhado, amarelo, cinza, são possíveis valores ou realizações da variável.

5.5.1.1 VARIÁVEIS QUALITATIVAS NOMINAIS

Surgem quando se definem categorias, tipos ou atributos, tais como sexo (masculino ou feminino), campo de estudo (engenharia, biologia, ciências humanas e ciências exatas), estado civil (solteiro, casado, divorciado, disquitado etc.), ou seja, são aquelas para as quais não existe nenhuma ordenação nas possíveis realizações, e conta-se o número de observações pertencentes a esta categoria. Exemplo: Tipo de raça bovina (holandês, nelore, gir, hereford, guzerá etc.).

5.5.1.2 VARIÁVEIS QUALITATIVAS ORDINAIS

São aquelas constituídas de valores relativos atribuídos para denotar ordem, ou hierarquia, isto é são aquelas para as quais existe certa ordem nos possíveis resultados, os possíveis valores ocupam uma ordem ou ranking dentro da série numérica. Como exemplo, tem-se o grau de instrução dos funcionários de uma agroindústria (1º grau, 2º grau, superior), dados oriundos de concurso de culinária, concurso de beleza, concurso de flores, concurso de cães, conjunto de notas atribuído a um restaurante (ótimo, bom, regular, ruim e péssimo), notas de resistência da cultura do milho a doença ferrugem (resistente, moderadamente resistente e susceptível). Exemplo: O grau de maturação de frutos de figo (verde, “de vez”, maduro).

5.5.2 VARIÁVEIS QUANTITATIVAS

São aquelas que apresentam como possíveis realizações valores resultantes de uma contagem ou mensuração, ou seja, os dados são de caráter numérico ou quando os resultados possuem uma estrutura numérica. Exemplo: O peso ao nascer em kg de um bezerro macho da raça bovina nelore. Essas variáveis podem ainda ser do tipo:

5.5.2.1 VARIÁVEIS QUANTITATIVAS DISCRETAS

“X” será uma variável discreta quando a menor diferença não-nula entre dois valores possíveis dessa variável for finita. Deve-se notar que teoricamente, X pode variar em todo o conjunto de números inteiros racionais, e não apenas inteiros positivos de zero a infinito (TOLEDO E OVALLE, 1987). Normalmente a variável discreta resulta de contagem, razão pela qual seus valores são expressos por meio de números inteiros não negativos, ou seja, o seu campo de definição é um conjunto finito ou infinito enumerável de valores. Como por exemplo,

pode-se citar o número de grãos por espiga de milho, o número de insetos por folha de planta etc. Exemplo: O número de vagens por planta na cultura do feijão Macassar: {0, 1, 2, 3, ...}.

5.5.2.2 VARIÁVEIS QUANTITATIVAS CONTÍNUAS

“X” é uma variável contínua, quando, ao passar de um valor real “a” para outro valor “b”, assume todos os valores intermediários entre a e b. Assim sendo, pode-se dizer que a variável contínua resulta geralmente de mensuração, e a escala numérica de seus possíveis valores corresponde ao conjunto \mathfrak{R} dos números reais (TOLEDO E OVALLE, 1987). São dados que podem assumir qualquer valor num intervalo contínuo, ou seja, num conjunto infinito não enumerável de valores, como por exemplo, dados obtidos por meio de mensuração, medição ou pesagem, tais como altura de plantas em centímetros, peso de frangos em gramas aos 30 dias de vida, velocidade do vento em m/s, temperatura corporal de um coelho durante um mês em graus Celsius, produtividade de uma cultura agrícola de algodão em toneladas por hectare, peso de 100 sementes de feijão macassar em gramas etc. EXEMPLO: A umidade relativa média do ambiente em porcentagem no dia 5 de janeiro de 2008 em Mossoró, RN, medida durante todo o dia.

5.6 CONSTANTE (K)

É uma característica populacional de um conjunto de indivíduos que toma apenas um único valor numérico de elemento para elemento. Exemplo: Número de ovos postos por 5 galinhas poedeiras: $K = \{10, 10, 10, 10, 10\}$.

5.7 DADO ESTATÍSTICO OBSERVADO OU OBSERVAÇÃO

É aquele conjunto de números, taxas e outras informações numéricas ou não, coletadas ou registradas em experimentos ou levantamentos amostrais associados a uma variável. Dentro desse contexto todo dado ou observação é um valor de uma variável. O valor observado será representado por letras minúsculas. Exemplos: x, y, z e w; x = 3 toneladas de algodão por hectare, w = 30 °C etc.

6. O MÉTODO ESTATÍSTICO: FASES DO TRABALHO ESTATÍSTICO

A realização de uma pesquisa deve passar, necessariamente pelas fases apresentadas abaixo (TOLEDO e OVALLE, 1987).

- i) Definição do problema: Saber exatamente o que se pretende pesquisar, ou seja, definir corretamente o problema.
- ii) Planejamento: determinar o procedimento necessário para resolver o problema, como levantar informações sobre o assunto objeto do estudo. É importante a escolha das perguntas em um questionário, que na medida do possível, devem ser fechadas. O levantamento de dados pode ser de dois tipos: Censitário e Amostragem. Outros elementos do planejamento de uma pesquisa são:
 - ii.1) Cronograma das atividades;
 - ii.2) Custos envolvidos;
 - ii.3) Exame das informações disponíveis;
 - ii.4) Delineamento da amostra.
- iii) Coleta de Dados: consiste na busca ou compilação dos dados. Pode ser classificado, quanto ao tempo em:
 - iii.1) Contínua (inflação, desemprego, etc);
 - iii.2) Periódica (Censo);
 - iii.3) Ocasional (pesquisa de mercado, pesquisa eleitoral)

- iv) Crítica dos dados: objetiva a eliminação de erros capazes de provocar futuros enganos. Faz-se uma revisão crítica dos dados suprimindo os valores estranhos ao levantamento.
- v) Apresentação dos dados: a organização dos dados denomina-se “Série Estatística”. Sua apresentação pode ocorrer por meio de tabelas e gráficos.
- vi) Análise e Interpretação dos Dados: consiste em tirar conclusões que auxiliem o pesquisador a resolver seu problema, descrevendo o fenômeno por meio do cálculo de medidas estatísticas, especialmente as de posição e as de dispersão.

7. SUMARIZAÇÃO DE DADOS

Essa é a parte que consiste em resumir os dados em Tabelas, gráficos ou medidas resumo, com o objetivo de melhor observar e estudar o comportamento do fenômeno estudado.

7.1. ANÁLISE EXPLORATÓRIA DE DADOS [AED]

O objetivo da análise exploratória de dados é examinar a estrutura interna dos dados e avaliar os relacionamentos sistemáticos entre muitas variáveis.

A análise exploratória de dados inclui um conjunto de ferramentas gráficas e descritivas, para explorar os dados, como pré-requisito para uma análise de dados mais formal (Predição e Testes de Hipóteses), e como parte integral formal da construção de modelos.

7.1.1. GRÁFICOS REPRESENTATIVOS DAS DISTRIBUIÇÕES DE FREQUÊNCIAS PARA VARIÁVEIS OU DADOS QUALITATIVOS

Os gráficos mais usados para representar as distribuições de frequências de dados qualitativos são os gráficos de Colunas, de Barras e o Gráfico Circular ou Setograma ou gráfico em “Pizza”.

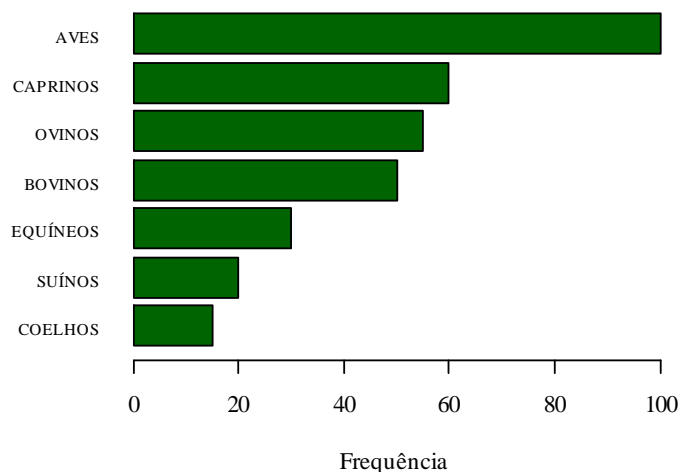
Vamos representar a distribuição de frequências mostrada anteriormente por meio desses três tipos de gráficos descritos.

i) GRÁFICO DE BARRAS

Os gráficos em barras têm por finalidade comparar, por meio de retângulos de igual largura e alturas proporcionais às respectivas grandezas. Cada barra representa a intensidade de uma modalidade do atributo.

As magnitudes das barras são representadas pelos respectivos comprimentos e seu traçado é feito tendo-se como referência uma escala horizontal. Em geral, as divisões da escala se prolongam em traços verticais por todo o gráfico, facilitando assim a leitura do comprimento de cada barra. A identificação da barra é inscrita à esquerda do gráfico. As barras só diferem em comprimento e não em largura, a qual é arbitrária, e elas devem vir separadas uma das outras pelo mesmo espaço, como regra prática pode-se tomar o espaço entre as barras como aproximadamente a metade ou dois terços de suas larguras.

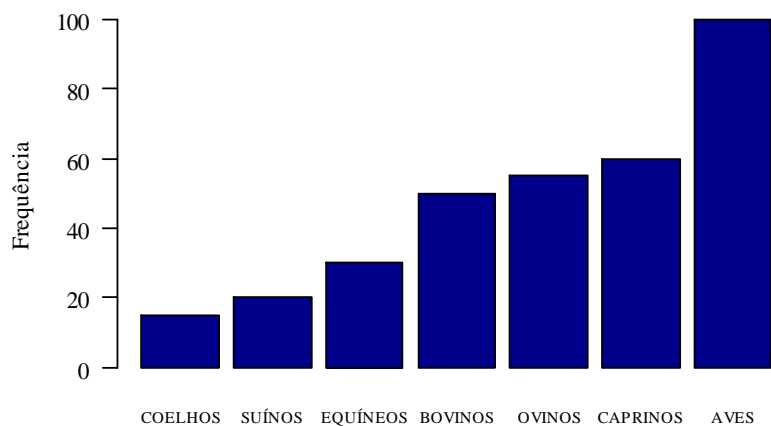
Exemplo: Gráfico de barras referente a distribuição de frequências relativas das espécies de animais exploradas em uma fazenda, em diversas propriedades agrícolas existentes no Estado do Paraná, cuja atividade principal é a pecuária.



ii) GRÁFICO DE COLUNAS

Os gráficos em colunas ou gráficos em barras verticais prestam-se à mesma finalidade que os gráficos em barras horizontais, sendo, entretanto, preferíveis a esses últimos quando as legendas a se inscreverem sob os retângulos forem breves. Caso contrário, o emprego do gráfico em barras é mais adequado. Todas as demais observações apresentadas anteriormente são extensíveis a esse tipo de gráfico.

Exemplo: Gráfico em colunas referente a distribuição de frequências relativas das espécies de animais explorados em uma fazenda, em diversas propriedades agrícolas existentes no Estado do Paraná, cuja atividade principal é a pecuária.



iii) GRÁFICO CIRCULAR OU EM SETORES OU SETOGRAMAS OU CARTOGRAMAS EM SETORES

Os gráficos em setores ou setogramas são usados para representar valores absolutos ou porcentagens complementares e pode ser feita por meio de diferentes tipos de gráficos.

Para construir esse tipo de gráfico parte-se do fato de que o número total de graus de um arco de circunferência é 360°. Assim, o número total de valores analisados (100%, se quisermos representar as porcentagens complementares) corresponderá a 360°. Cada uma das parcelas componentes do total dos valores poderá, então, ser expressa em graus, e a correspondência se fará por meio de uma regra de três simples.

Por exemplo, suponha que n é o número total de dados e n_i é o número de dados de uma dada categoria. Então pode-se lançar mão da regra de três:

$$n \rightarrow 360 \text{ graus}$$
$$n_i \rightarrow x_i = \frac{360 \times n_i}{n} \text{ graus}$$

O gráfico pode ser feito a partir de um círculo de circunferência (360°) como já comentado anteriormente ou de um semicírculo (180°) ou de um quadrante (90°) conforme se utilize o semicírculo ou o quadrante para a representação completa do fenômeno.

Exemplo: Gráfico de setores referente à distribuição de frequências relativas das espécies de animais explorados em uma fazenda, em diversas propriedades agrícolas existentes no Estado do Paraná, cuja atividade principal é a pecuária.

