

Ciência da Computação

Estatística Básica

Aula 2 – Organização de dados em tabelas e séries. Tipo de séries. Dados absolutos X Relativos em Estatística.

Professora Ma. Tainara Volan
tainaravolan@gmail.com

Fases do método estatístico

a. Definição do problema:

O que se pretende investigar, pesquisar, avaliar? É quando se define qual é o problema que quer se resolver, delimitando-se o objeto de estudo de forma viável em relação ao tempo e aos recursos disponíveis.

b. Planejamento:

Após a definição do problema a ser estudado, o planejamento consiste em determinar os procedimentos para a investigação e solução do problema. É a etapa onde são estabelecidos os detalhes mais importantes do estudo: o cronograma geral, a metodologia da coleta de dados, a definição do tamanho da amostra, entre outros.

c. Coleta ou levantamento dos dados:

É a obtenção, reunião e registro sistemático de dados.

Fases do método estatístico

d. Crítica dos dados:

Etapa onde são observadas as discrepâncias nos dados obtidos e, se necessário, decidir se haverá uma nova coleta de dados, descarte do dado discrepante ou complementação das informações sobre esse dado.

e. Apuração dos dados ou sumarização:

É o processo de sumarização dos dados obtidos mediante critério de classificação.

f. Apresentação dos dados:

Após a apuração, os dados são organizados conforme o objetivo e apresentados sob a forma de tabelas, gráficos e medidas.

g. Análise e interpretação dos dados:

É a etapa onde são feitas medidas complementares a partir dos dados coletados que darão suporte a deduções e/ou induções sobre as informações obtidas, como também as conclusões sobre o estudo em questão. É o resultado do trabalho estatístico, com solução para o problema definido na primeira fase. Exige conhecimento técnico do processo, fenômeno ou evento que se está analisando.

Estatística Descritiva

Após a etapa de coleta dos dados, normalmente temos um conjunto extenso de valores e informações que precisam ser **ordenados e organizados** de tal forma que possamos ter uma **visão global do fenômeno analisado**. Representar o conjunto de valores por meio de tabelas ou gráficos adequados irá permitir uma boa caracterização das informações que temos, com as quais poderemos realizar diagnósticos e conclusões ou, ainda, fazer comparações com outros conjuntos semelhantes de dados.

Série Estatística

Uma **série estatística** consiste em um conjunto de dados ordenado segundo uma característica comum, ou seja, os dados referem-se a uma mesma variável. Uma série estatística, comumente, é representada através de uma tabela ou de um gráfico, conforme melhor ficar representado o conjunto de dados que queremos analisar.

Tabelas

Os dados depois de tratados podem ser apresentados em tabelas. Existem normas nacionais para a organização de tabelas, ditadas pela ABNT. Essas normas não serão tratadas aqui, mas convém saber que as tabelas devem ter os seguintes componentes:

- Título: Precede a tabela e explica, em poucas palavras, o dado em estudo. Se for o caso, indica o tempo e o lugar a que os dados se referem.
- Cabeçalho: Especifica o conteúdo de cada coluna.
- Coluna Indicadora: Especifica em cada linha os valores que os dados podem assumir.
- Corpo da tabela: Apresenta a frequência dos dados.
- Fonte: Especifica a entidade, o pesquisador ou pesquisadores que forneceram os dados, quando esses não foram coletados por você.

Tabelas

Exemplo

Título	Valor em dólares dos principais produtos que o Brasil vende à Argentina	
Cabeçalho	Produto	Valor em dólares (em bilhões)
Coluna Indicadora	Automóveis	606
	Veículos de carga	541
	Autopeças	531
	Motores	264
	Minério	248
	Tratores	130
Fonte: Época 25 de janeiro de 1999		

Tabelas

Considerações:

- O título da tabela deve indicar a natureza e a abrangência geográfica e/ou temporal dos dados. É colocado na parte superior, precedido da palavra Tabela e de seu número de ordem.
- As tabelas são numeradas consecutivamente e independentemente das ilustrações, em algarismos arábicos. A numeração pode ser subordinada ou não a capítulos ou seções de um documento.
- A tabela não deve ser fechada lateralmente (sem fios laterais);
- Não há obrigatoriedade de linha (fio) vertical entre as colunas, mas esta pode ser utilizada desde que seja necessário o que ocorre quando a tabela apresenta muita informação (muitas colunas e/ou muitas linhas);
- Não devem ser utilizados traços (fios) horizontais separando as linhas com exceção do cabeçalho e da última linha;
- Nenhuma célula deve ficar em branco; a ausência do dado é expressa por um traço (-) e a falta de conhecimento deste (dado ignorado) é expressa por três pontos (...);

Tabelas

Considerações:

- Quando há dúvida quanto a um fato numérico, pode-se ainda segui-lo de um ponto de interrogação (?).
- Notas e chamadas são utilizadas para clarificar os dados. As notas fornecem informações de natureza geral, destinadas a explicitar ou a esclarecer o conteúdo da tabela ou a indicar a metodologia adotada no levantamento de dados, enquanto as chamadas se referem a informações específicas. Ambas são colocadas no rodapé da tabela, abaixo da fonte, sendo as notas listadas assim: 1, 2, 3 etc, e as chamadas (1), (2), (3) etc. Seu emprego deve ser evitado ao máximo, dado que contrariam o princípio de síntese proposto na elaboração de tabelas e gráficos; quando absolutamente necessárias devem ser redigidas de maneira muito concisa, indicando claramente os dados da tabela a que se referem.

Tabelas

Considerações:

- Recomenda-se a citação da fonte quando reproduzidas de outros documentos. A prévia autorização do autor se faz necessária, não sendo mencionada na mesma. Quando os dados apresentados na tabela foram levantados pelo autor do trabalho por meio de uma pesquisa de campo (questionários, formulários, entrevistas), pode-se utilizar como fonte as expressões o autor ou pesquisa de campo;
- As tabelas devem ser inseridas o mais próximo possível do trecho a que se referem;
- Quando uma tabela, por excessiva altura, tiver de ocupar mais de uma página, não deve ser delimitada na parte inferior, repetindo-se o cabeçalho na página seguinte. Neste caso, deve-se usar no alto do cabeçalho ou dentro da coluna indicadora a designação Continua ou Conclusão, conforme o caso;

Tabelas

Tabela 2.2 - Utilização de medida anticoncepcional de acordo com o estado civil – USU – 1995

Medida	Total	Estado Civil	
		Solteiro	Casado
Pílula	18	12	6
Tabela	10	9	1
Ligadura	10	2	8
Coito Interrompido	5	5	-
Diu	3	-	3
Preservativo	3	3	-
Diafragma	1	-	1
Total	50	31	19

Fonte: WILMER, C., CASTELLO, G. & DE FARIAS, A.

Tabela 2.3. Taxas anuais de crescimento da produção industrial por categoria de uso - 2003-2004

Produção Industrial	2003	2004
Indústria Geral	0	8,3
Bens de Capital	2,2	19,7
Bens Intermediários	2	7,4
Bens de Consumo	-2,7	7,3
Bens de Consumo Durável	3	21,8
Bens de Consumo Não Durável	-3,9	4

Fonte: IBGE, Diretoria de pesquisas, Coordenação de Indústria. Pesquisa Industrial Mensal.

Classificações das Séries Estatísticas

As séries estatísticas ou tabelas são classificadas de acordo com o conteúdo apresentado. Utiliza-se como base de classificação, a coluna indicadora.

Séries históricas, temporais, cronológicas ou marchas

As séries temporais ou cronológicas resultam tabelas que apresentam os valores da variável em estudo, em determinado local, discriminados segundo intervalos de tempo variáveis.

Taxa de escolarização das pessoas de 18 e 19 anos de idade – Brasil – 2001/2008

Ano	Taxa de escolarização
2001	51,4
2002	51,1
2003	51,7
2004	48,5
2005	47,6
2006	47,0
2007	45,0
2008	46,0

Fonte: IBGE, 2010.

**Tabela 2.4 – Preço do acém no varejo
– São Paulo – 1989-1992**

Anos	Preço médio (US\$)
1989	2,24
1990	2,73
1991	2,12
1992	1,89

Fonte: APA.

Séries geográficas, espaciais, territoriais ou de localização

As tabelas com séries geográficas ou espaciais apresentam os valores da variável em estudo, em determinado tempo, discriminados segundo regiões.

Taxa de escolarização das pessoas de 18 e 19 anos de idade, por regiões – Brasil –2008	
Regiões	Taxa de escolarização
Norte	51,1
Nordeste	50,6
Centro-Oeste	47,1
Sudeste	42,9
Sul	41,4

Fonte: IBGE, 2010.

Tabela 2.5 – Duração média dos estudos superiores - 1994

Países	Número médio de anos
Itália	7,5
Alemanha	7,0
França	7,0
Holanda	5,9
Inglaterra	Menos de 4

FONTE: Revista Veja.

Séries mistas, dupla entrada ou conjugadas

São tabelas que apresentam a variação de valores de mais de uma variável, conjugando duas ou mais séries. Quando duas séries são conjugadas em uma mesma tabela temos uma tabela de dupla entrada, onde aparecem duas ordens de classificação: uma na linha (horizontal) e outra na coluna (vertical). Por apresentar mais de uma característica dos dados ao mesmo tempo, este tipo de tabela exige sempre mais de duas colunas.

Produção de Carne no Brasil, 1970-2000				
Origem	Produção de carne (milhões de toneladas)			
	1970	1980	1990	2000
Bovinos	1.845	2.850	4.115	6.579
Suínos	767	980	1.050	2.600
Frangos	366	1.370	2.356	5.980

Fonte: MMA, 2006.

**Tabela 2.7 – Balança comercial do Brasil.
1989 – 1993.**

Especifica- ções	Valor (Us\$ 1 000 000)				
	1989	1990	1991	1992	1993
Exportação	34383	31414	31620	35793	38783
Importação	18263	20661	21041	20554	25711

Fonte: Ministério da Fazenda

Quadros

Inicialmente os dados dos questionários ou de qualquer outra forma de coleta devem ser organizados em Quadros, em que são listadas as informações obtidas para cada participante. O quadro é completamente fechado e não se propõe a resumir os dados. De acordo com a NBR 14724 (2011), a identificação deve ser feita na parte superior precedida da palavra Quadro, seguida do número de ordem no texto, em algarismos arábicos e respectivo título. Após o quadro insere-se a fonte, mesmo quando for do próprio autor.

Quadro 2.1 –Características de estudante da Universidade Santa Úrsula que utilizam Métodos Anticoncepcionais.

Sujeito	Sexo	Idade	Estado Civil	Nível de Instrução	Quem toma a medida	Qual a medida tomada
1	F	28	C	Superior	Mulher	Pílula
2	M	29	S	2º grau	Mulher	Tabela
3	f	36	S	Superior	Mulher	Ligadura
...*						
47	F	42	C	Superior	Mulher	Ligadura
48	F	30	C	2º grau	Mulher	Ligadura
49	F	46	C	1º grau	Mulher	Ligadura
50	F	29	C	2º grau	Mulher	Pílula

* Foi feito um corte no Quadro. O quadro completo pode ser encontrado na fonte abaixo.

Fonte:BUNCHAFT&KELLNER, *Estatística sem Mistérios*, Ed. Vozes, Petrópolis: 1997, pág. 31 –34.

Dados Absolutos e Dados Relativos

Os dados estatísticos resultantes da coleta direta da fonte, sem outra manipulação senão a contagem ou medida são chamados **dados absolutos**.

A leitura dos dados absolutos é sempre inexpressiva; embora esses dados traduzam um resultado exato e fiel, não tem a virtude de ressaltar de imediato as suas conclusões numéricas. Daí o uso imprescindível que faz a Estatística dos dados relativos.

Dados relativos: é o resultado de comparações por quociente (razões) que se estabelecem entre dados absolutos e tem por finalidade realçar ou facilitar as comparações entre quantidades.

Traduzem-se os dados relativos, em geral, por meio de percentagens, índices, coeficientes e taxas. Consideremos

Dados Absolutos e Dados Relativos

Tabela 1 – Matrículas nas escolas da cidade A em 2003

Categorias	n. de alunos	Por 1	%
1º grau	19.286	0,91	91
2º grau	1.681	0,08	8
3º grau	234	0,01	1
Total	21.201	1,0000	100,00

Fonte: dados fictícios

Dados Absolutos e Dados Relativos

Tabela 2 – Matrículas nas escolas da cidade A e B em 2003

	Cidade A		Cidade B	
Categorias	n. de alunos	%	n. de alunos	%
1º grau	19.286	91	38.660	91
2º grau	1.681	8	3.399	8
3º grau	234	1	424	1
Total	21.201	100,0	42.483	100,0

Fonte: dados fictícios

Qual das cidades tem, comparativamente, maior número de alunos em cada grau?

Exercícios

1) Complete a tabela a 2.16:

Tabela 2.16 – Alunos matriculados em escolas fictícias – 1995.

Escolas	Nº de alunos	Dados relativos	
		Por 1	Por 100
A	175	0,098	9,8
B	222		
C	202		
D	362		
E	280		
F	540	1,000	100,0
TOTAL	1781		

Dados fictícios.

Exercícios

2) Uma escola apresentava, no final do ano, o seguinte quadro:

Tabela 2.17 – Matrículas na escola XYZ – 1999.

SÉRIES	Matrículas		
	Março	Novembro	%
1 ^a	480	475	
2 ^a	458	456	
3 ^a	436	430	
4 ^a	420	420	
Total	1794	1781	

Fonte: Dados fictícios

- a) Calcule a taxa de evasão por série.
- b) Calcule a taxa de evasão da escola.

Exercícios

3) Para melhorar o uso de um determinado aplicativo, o programador fez uma análise da quantidade de estrelas que 100 usuários (escolhidos aleatoriamente) deram ao serviço:

Estrelas	Frequência absoluta
1 estrela	20
2 estrelas	8
3 estrelas	10
4 estrelas	7
5 estrelas	55
Total	100

Analizando os dados encontrados, é correto afirmar que:

- A) Mais da metade dos usuários deu, no máximo, 3 estrelas para o aplicativo.
- B) Menos da metade dos usuários deu 5 estrelas para o aplicativo.
- C) Um quarto dos usuários deu notas entre 2 e 4 estrelas.
- D) Exatamente 38 usuários deram, pelo menos, 3 estrelas

Exercícios

4) Considere a tabela com os salários dos funcionários de uma empresa a seguir:

Salário	Frequência absoluta
Menor que R\$ 1000	6
Maior que R\$ 1000 e menor que R\$ 1500	8
Maior que R\$ 1500 e menor que R\$ 2000	7
Maior que R\$ 2000	2
Total	25

Então, podemos afirmar que a frequência absoluta dos funcionários que ganham, pelo menos, R\$1500 é igual a: