

ESTATÍSTICA BÁSICA

Por quê?

Em alguma fase de seu trabalho, o pesquisador se vê às voltas com o problema de analisar e entender um conjunto de dados. Se forem informações sobre uma amostra ou população, ele necessitará resumir os dados para que eles sejam informativos, ou para compará-los com outros resultados, ou ainda para julgar sua adequação a alguma teoria

ESSÊNCIA DA CIÊNCIA

De modo geral podemos dizer que a essência da ciência é a *observação* e que seu objetivo básico é a *Inferência*. A inferência pode ser *Indutiva* (do específico ao geral) ou *Dedutiva* (**das premissas às conclusões**)

A *inferência estatística* é uma das etapas da estatística. É a parte da metodologia científica que tem por objetivos a coleta, redução, análise e modelagem dos dados.

Nesta aula vamos dar enfoque à análise exploratória dos dados.

CONCEITOS

POPULAÇÃO

Ao coletar os dados referentes às características de um grupo de objetos ou indivíduos, tais como as alturas e pesos dos estudantes de uma universidade ou os números de parafusos defeituosos, é muitas vezes impossível ou impraticável observar todo o grupo, especialmente se for muito grande. Todo o grupo ou todas as realizações possíveis de uma variável qualquer é denominado população ou universo. Uma população pode ser finita ou infinita.

CONCEITOS

AMOSTRA

Como em muitos casos é impraticável observar a população, recorre-se ao artifício de se coletar um conjunto de realizações que sejam representativas da população. Este conjunto de realizações é denominado amostra. Se a amostra é representativa, importantes conclusões sobre a população podem ser inferidas de sua análise. A parte da estatística que trata das condições sob as quais essas inferências são válidas chama-se estatística indutiva ou inferência. Como essa inferência não pode ser absolutamente certa, a linguagem da probabilidade é muitas vezes usada, no estabelecimento das conclusões.

DADOS BRUTOS: Dados brutos são aqueles que ainda não foram numericamente organizados.

TIPOS DE VARIÁVEIS

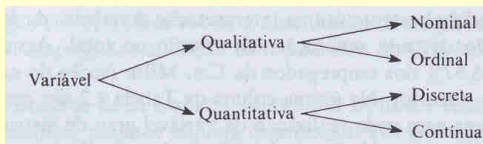
Uma variável é um símbolo, como X, Y, H que pode assumir qualquer um de um conjunto de valores que lhe são atribuídos, conjunto este chamado domínio da variável. Uma variável é denominada **Constante** quando esta assume apenas um valor.

Uma variável é considerada **Qualitativa** quando apresenta como possíveis realizações uma qualidade (ou atributo) do indivíduo pesquisado. Ex: sexo, educação, estado civil. Por outro lado, uma variável é considerada **Quantitativa** quando apresentam como possíveis realizações, números resultantes de uma contagem ou mensuração. Ex: número de filhos de um casal e altura e peso de alunos, respectivamente.

Uma variável é tida como quantitativa **Contínua** quando esta pode assumir teoricamente qualquer valor entre dois dados. Caso contrário é denominada quantitativa **Discreta**. De modo geral, as **medições** dão origem a **dados contínuos**, enquanto que as **enumerações** e contagens resultam em **dados discretos**.

CLASSIFICAÇÃO DAS VARIÁVEIS

Uma variável é denominada qualitativa **Nominal** quando não existe qualquer possibilidade de ordenação nas possíveis realizações. É considerada qualitativa **Ordinal** quando existe alguma ordem nos possíveis resultados.



N.º	Estado civil	Grau de instrução	N.º de filhos	Salário (R\$ Sal. Mês)	Idade anos meses	Região de procedência
1	solteiro	1.º grau	—	4,00	26 03	Interior
2	casado	1.º grau	1	4,56	32 10	Capital
3	casado	1.º grau	2	5,25	36 05	Capital
4	solteiro	2.º grau	—	5,73	20 10	Outro
5	solteiro	1.º grau	—	6,26	40 07	Outro
6	solteiro	1.º grau	0	6,66	28 00	Interior
7	solteiro	1.º grau	—	6,86	41 00	Interior
8	solteiro	1.º grau	—	7,39	43 04	Capital
9	casado	2.º grau	1	7,59	34 10	Capital
10	solteiro	2.º grau	—	7,44	24 06	Capital
11	casado	2.º grau	2	8,12	33 06	Interior
12	solteiro	1.º grau	—	8,46	27 11	Capital
13	solteiro	2.º grau	—	8,74	37 05	Outro
14	casado	1.º grau	3	8,95	44 02	Capital
15	casado	2.º grau	0	9,13	30 05	Interior
16	casado	2.º grau	2	9,35	38 08	Outro
17	casado	2.º grau	1	9,77	31 07	Capital
18	casado	1.º grau	2	9,86	39 07	Outro
19	solteiro	superior	—	10,53	25 08	Interior
20	solteiro	2.º grau	—	10,76	37 04	Interior
21	casado	2.º grau	1	11,06	30 09	Outro
22	solteiro	2.º grau	—	11,59	34 02	Capital
23	solteiro	1.º grau	—	12,00	41 00	Outro
24	casado	superior	0	12,79	26 01	Outro
25	casado	2.º grau	2	13,23	32 05	Interior
26	casado	2.º grau	2	13,66	35 06	Capital
27	solteiro	1.º grau	0	13,85	46 07	Outro
28	casado	2.º grau	2	14,69	29 08	Interior
29	casado	2.º grau	5	14,71	40 06	Interior
30	casado	2.º grau	2	15,99	35 10	Capital
31	solteiro	superior	—	16,22	31 05	Outro
32	casado	2.º grau	1	16,61	36 04	Capital
33	casado	superior	3	17,26	43 07	Capital
34	solteiro	superior	—	18,75	33 07	Capital
35	casado	2.º grau	2	19,40	48 11	Capital
36	casado	superior	3	23,96	42 02	Interior

Dados Qualitativo ordinal (Estado civil, Grau de instrução, Região de procedência)

Dados Qualitativo nominal (Região de procedência)

Dados Quantitativo discreto (N.º de filhos)

Dados Quantitativo contínuo (Salário, Idade)

DISTRIBUIÇÃO DE FREQUÊNCIA

Quando se estuda uma variável, o maior interesse do pesquisador é conhecer a distribuição dessa variável através das possíveis realizações (valores) da mesma. A distribuição de frequência é uma boa maneira de se dispor um conjunto de valores, de modo a se ter uma boa idéia global sobre esses valores, ou seja, de sua distribuição.

DISTRIBUIÇÃO DE FREQUÊNCIA

Quando se resume grandes quantidades de dados brutos, costuma-se freqüentemente distribuí-los em classes ou categorias e determinar o número de indivíduos pertencentes a cada uma das classes, denominado *freqüência da classe*. Um arranjo tabular dos dados por classes, juntamente com as freqüências correspondentes, é denominado *distribuição de freqüência* ou *tabela de freqüência*.

HISTOGRAMA: é uma representação gráfica de uma distribuição de freqüência por meio de retângulos justapostos.

TABELA DE FREQUÊNCIA

N.º	Estado civil	Grau de instrução	N.º de filhos	Salário (R\$ Sal. Mês)	Idade anos meses	Região de procedência	Gr. de instrução	Contagem	Freqüência n_i	Proporção f_i	Porcentagem $100 \cdot f_i$
1	solteiro	1.º grau	—	4,00	26 03	Interior	1.º grau	HHH+HHH	12	0,3333	33,33
2	casado	1.º grau	1	4,56	32 10	Capital	2.º grau	HHH+HHH+HHH	18	0,5000	50,00
3	casado	1.º grau	2	5,25	36 05	Capital	Superior	HHH+	5	0,1667	16,67
4	solteiro	2.º grau	—	5,73	20 10	Outro					
5	solteiro	1.º grau	—	6,26	40 07	Outro					
6	casado	1.º grau	0	6,66	28 00	Interior					
7	solteiro	1.º grau	—	6,86	41 00	Interior					
8	solteiro	1.º grau	—	7,39	43 04	Capital					
9	casado	2.º grau	1	7,59	34 10	Capital					
10	solteiro	2.º grau	—	7,44	23 06	Capital					
11	casado	2.º grau	2	8,12	33 06	Interior					
12	solteiro	1.º grau	—	8,46	27 11	Capital					
13	solteiro	2.º grau	—	8,74	37 05	Outro					
14	casado	1.º grau	3	8,95	44 02	Outro					
15	casado	2.º grau	0	9,13	30 05	Interior					
16	solteiro	2.º grau	—	9,35	38 08	Outro					
17	casado	2.º grau	1	9,77	31 07	Capital					
18	casado	1.º grau	2	9,86	39 07	Outro					
19	solteiro	superior	—	10,53	25 08	Interior					
20	solteiro	2.º grau	—	10,76	37 04	Interior					
21	casado	2.º grau	1	11,06	30 09	Outro					
22	solteiro	2.º grau	—	11,59	34 02	Capital					
23	solteiro	1.º grau	—	12,00	41 00	Outro					
24	casado	superior	0	12,79	26 01	Outro					
25	casado	2.º grau	2	13,23	32 05	Interior					
26	casado	2.º grau	2	13,66	35 06	Capital					
27	solteiro	1.º grau	0	13,85	46 07	Outro					
28	casado	2.º grau	2	14,69	29 08	Interior					
29	casado	2.º grau	5	14,71	40 06	Interior					
30	casado	2.º grau	2	15,99	35 10	Capital					
31	solteiro	superior	—	16,22	31 05	Outro					
32	casado	2.º grau	1	16,61	36 04	Capital					
33	casado	superior	3	17,26	43 07	Capital					
34	solteiro	superior	—	18,75	33 07	Capital					
35	casado	2.º grau	2	19,40	48 11	Capital					
36	casado	superior	3	23,96	42 02	Interior					
TOTAL									36	1,0000	100,00

Classe de salários	Freqüência n_i	Porcentagem acumulada $100 \cdot F_i$
4,00 — 8,00	10	27,78
8,00 — 12,00	12	33,33
12,00 — 16,00	8	33,33
16,00 — 20,00	5	33,33
20,00 — 24,00	1	33,33
TOTAL	36	100,00

Gr. de instrução	Freqüência	Porcentagem
1.º grau	650	32,50
2.º grau	1.020	51,00
Superior	330	16,50
TOTAL	2.000	100,00

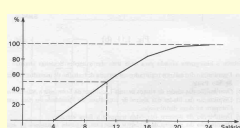
FREQUÊNCIA ACUMULADA

Classe de salários	Freqüência n_i	Freqüência acumulada N_i	Porcentagem acumulada $100 \cdot F_i$
4,00 — 8,00	10	10	27,78
8,00 — 12,00	12	22	33,33
12,00 — 16,00	8	30	33,33
16,00 — 20,00	5	35	33,33
20,00 — 24,00	1	36	33,33
TOTAL	36	—	100,00

A freqüência acumulada indica quantos elementos, ou que porcentagem deles, estão abaixo de um certo valor.

Assim, observando a tabela ao lado pode-se afirmar que, por exemplo, 27,78% dos indivíduos ganham até 8 salários e 97,22% ganham até 20 salários.

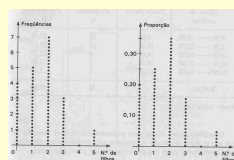
O gráfico pode fornecer informações adicionais como: qual salário s , tal que 50% dos funcionários ganham menos do que s .



REPRESENTAÇÃO GRÁFICA DE DADOS DISCRETOS

N.º	Estado civil	Grau de instrução	N.º de filhos	Salário (R\$ Sal. Mês.)	Idade atual meses	Região de procedência
1	solteiro	1º grau	0	4,00	26	Interior
2	casado	1º grau	1	4,50	32	Capital
3	casado	1º grau	3	5,20	30	Capital
4	solteiro	2º grau	0	5,70	30	Outro
5	solteiro	1º grau	0	6,20	40	Interior
6	casado	1º grau	6	6,60	28	Interior
7	casado	1º grau	6	6,80	41	Interior
8	solteiro	1º grau	7	7,30	43	Capital
9	casado	2º grau	7	7,50	34	Capital
10	solteiro	2º grau	1	7,80	23	Interior
11	casado	2º grau	2	8,10	33	Interior
12	solteiro	1º grau	2	8,40	27	Capital
13	solteiro	2º grau	0	8,70	37	Interior
14	casado	1º grau	3	8,90	44	Outro
15	casado	2º grau	3	9,10	30	Interior
16	solteiro	2º grau	3	9,30	38	Outro
17	casado	2º grau	1	9,70	31	Capital
18	casado	1º grau	2	9,90	31	Outro
19	solteiro superior	superior	0	10,30	23	Interior
20	solteiro	2º grau	0	10,70	37	Interior
21	casado	2º grau	0	11,00	30	Outro
22	solteiro	2º grau	1	11,50	34	Capital
23	solteiro	1º grau	0	12,00	41	Outro
24	casado superior	superior	0	12,50	28	Interior
25	casado	2º grau	0	13,20	32	Interior
26	casado	2º grau	0	13,60	35	Outro
27	solteiro	1º grau	0	13,85	46	Outro
28	casado	2º grau	0	14,00	29	Interior
29	casado	2º grau	0	14,70	40	Interior
30	casado	2º grau	0	15,00	33	Capital
31	solteiro superior	superior	0	16,20	31	Outro
32	casado	2º grau	1	16,60	36	Interior
33	casado superior	superior	3	17,20	43	Capital
34	solteiro superior	superior	0	18,70	33	Capital
35	casado	2º grau	2	19,40	48	Capital
36	casado superior	superior	0	21,30	42	Interior

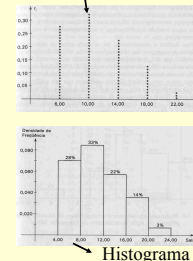
N.º de filhos x_i	Frequência n_i	Porcentagem $100 \cdot f_i$
0	4	20%
1	5	25%
2	7	35%
3	3	15%
5	1	5%
TOTAL	20	100%



REPRESENTAÇÃO GRÁFICA DE DADOS CONTÍNUOS

N.º	Estado civil	Grau de instrução	N.º de filhos	Salário (R\$ Sal. Mês.)	Idade atual meses	Região de procedência
1	solteiro	1º grau	0	4,00	26	Interior
2	casado	1º grau	1	4,50	32	Capital
3	casado	1º grau	2	5,20	30	Capital
4	solteiro	2º grau	0	5,70	30	Outro
5	solteiro	1º grau	0	6,20	40	Interior
6	casado	1º grau	6	6,60	28	Interior
7	solteiro	1º grau	0	6,80	41	Interior
8	solteiro	1º grau	0	7,30	43	Capital
9	casado	2º grau	1	7,50	34	Interior
10	solteiro	2º grau	0	7,80	23	Outro
11	casado	2º grau	2	8,10	33	Interior
12	solteiro	1º grau	2	8,40	27	Capital
13	solteiro	2º grau	0	8,70	37	Interior
14	casado	1º grau	3	8,90	44	Outro
15	casado	2º grau	3	9,10	30	Interior
16	solteiro	2º grau	3	9,30	38	Outro
17	casado	2º grau	1	9,70	31	Capital
18	casado	1º grau	2	9,90	31	Outro
19	solteiro superior	superior	0	10,30	23	Interior
20	solteiro	2º grau	0	10,70	37	Interior
21	casado	2º grau	0	11,00	30	Outro
22	solteiro	2º grau	1	11,50	34	Capital
23	casado	2º grau	0	12,00	41	Outro
24	casado superior	superior	0	12,50	28	Interior
25	casado	2º grau	2	13,20	32	Interior
26	casado	2º grau	2	13,60	35	Outro
27	solteiro	1º grau	0	13,85	46	Outro
28	casado	2º grau	0	14,00	29	Interior
29	casado	2º grau	0	14,70	40	Interior
30	casado	2º grau	0	15,00	33	Capital
31	solteiro superior	superior	0	16,20	31	Outro
32	casado	2º grau	1	16,60	36	Interior
33	casado superior	superior	3	17,20	43	Capital
34	solteiro superior	superior	0	18,70	33	Capital
35	casado	2º grau	2	19,40	48	Capital
36	casado superior	superior	0	21,30	42	Interior

Classe de salário	Ponto médio	Frequência n_i	Porcentagem $100 \cdot f_i$
4,00 — 8,00	6,00	10	20%
8,00 — 12,00	10,00	12	24%
12,00 — 16,00	14,00	8	16%
16,00 — 20,00	18,00	5	10%
20,00 — 24,00	22,00	1	2%
TOTAL		36	100,00



NÚMERO DE CLASSES DA DISTRIBUIÇÃO DE FREQUÊNCIA PARA DADOS AGRUPADOS

Tamanho da amostra (n)	Número de Classes (k)
Até 100	\sqrt{n} (inteiro mais próximo)
Acima de 100	$5 \log_{10} n$ (inteiro mais próximo)

Amplitude total e de classes

Amplitude total (R)

R = Valor máximo – valor mínimo

Amplitude de classes (h)

h = R/k

Exercícios

Distribuição de frequência

33	35	35	39	41	41	42	45	47	48
50	52	53	54	55	55	57	59	60	60
61	64	65	65	66	66	66	66	67	68
69	71	73	73	74	74	76	77	77	78
80	81	84	85	85	88	89	91	94	97

Pede-se:

- 1) Amplitude total, 2) Numero de classes, 3) frequência absoluta, 4) frequência relativa, 5) frequência acumulada e 6) o histograma.