

Conceitos Básicos II

Visualização de dados

Distribuição de frequências

Vejamos o exemplo ao lado.

- É uma tabela com dados dos funcionários de uma empresa:

1. Estado civil;
2. Grau de instrução;
3. Nº de filhos
4. Salários;
5. Idade
6. Região de procedência

N	Estado Civil	Grau de Instrução	N de Filhos	Salario (x Sal Min)	Idade (em anos)	Meses	Região de Procedência
1	solteiro	ensino fundamental		4,00	26	3	interior
2	casado	ensino fundamental	1	4,56	32	10	capital
3	casado	ensino fundamental	2	5,25	36	5	capital
4	solteiro	ensino médio		5,73	20	10	outra
5	solteiro	ensino fundamental		6,26	40	7	outra
6	casado	ensino fundamental	0	6,66	28	0	interior
7	solteiro	ensino fundamental		6,86	41	0	interior
8	solteiro	ensino fundamental		7,39	43	4	capital
9	casado	ensino médio	1	7,59	34	10	capital
10	solteiro	ensino médio		7,44	23	6	outra
11	casado	ensino médio	2	8,12	33	6	interior
12	solteiro	ensino fundamental		8,46	27	11	capital
13	solteiro	ensino médio		8,74	37	5	outra
14	casado	ensino fundamental	3	8,95	44	2	outra
15	casado	ensino médio	0	9,13	30	5	interior
16	solteiro	ensino médio		9,35	38	8	outra
17	casado	ensino médio	1	9,77	31	7	capital
18	casado	ensino fundamental	2	9,80	39	7	outra
19	solteiro	superior		10,53	25	8	interior
20	solteiro	ensino médio		10,76	37	4	interior
21	casado	ensino médio	1	11,06	30	9	outra
22	solteiro	ensino médio		11,59	34	2	capital
23	solteiro	ensino fundamental		12,00	41	0	outra
24	casado	superior	0	12,79	26	1	outra
25	casado	ensino médio	2	13,23	32	5	interior
26	casado	ensino médio	2	13,60	35	0	outra
27	solteiro	ensino fundamental		13,85	46	7	outra
28	casado	ensino médio	0	14,69	29	8	interior
29	casado	ensino médio	5	14,71	40	6	interior
30	casado	ensino médio	2	15,99	35	10	capital
31	solteiro	superior		16,22	31	5	outra
32	casado	ensino médio	1	16,61	36	4	interior
33	casado	superior	3	17,26	43	7	capital
34	solteiro	superior		18,75	33	7	capital
35	casado	ensino médio	2	19,40	48	11	capital
36	casado	superior	3	23,30	42	2	interior

Distribuição de frequências

- Dificuldade em estudar grande quantidade de dados;
- Tabelas de frequências possibilitam o resumo dos dados, de forma a facilitar a visualização;
- Para tanto, as variáveis são agrupadas de acordo com suas ocorrências;
- Embora útil, o processo de construção de distribuição de frequências pode causar perda de informações.

Distribuição de frequências

Regra Prática

1. Identificar o maior e o menor valor dos dados brutos;
2. Calcular a amplitude total dos dados;
3. Dividir a amplitude total dos dados para que todas as classes tenham a mesma amplitude de classe;
4. Agrupar os dados dentro das respectivas classes formadas.

Distribuição de frequências

Regra Prática

OBSERVAÇÕES IMPORTANTES

1. O número de intervalos de classe é, geralmente tomado entre 5 a 20, dependendo dos dados;
2. Os intervalos de classe são, preferencialmente escolhidos, de forma que o ponto médio do intervalo coincidam com dados observados;

Distribuição de frequência

Voltando ao exemplo...

Retornando a tabela apresentada anteriormente, vamos agrupar os dados de acordo com o grau de instrução:

Grau de instrução	Frequência n_i	Proporção f_i	Porcentagem $100 f_i$
Fundamental	12	0,3333	33,33
Médio	18	0,5000	50,00
Superior	6	0,1667	16,67
Total	36	1,0000	100,00

Frequência: número de observações;

Proporção ou Frequência Relativa: número de observações dividido pelo total;

Distribuição de frequências

Um problema...

Seria possível comparar essas informações da tabela (grau de instrução) com a tabela abaixo?

Grau de instrução	Frequência n_i	Porcentagem $100 f_i$
Fundamental	1.650	32,50
Médio	1.020	51,00
Superior	1.330	16,50
Total	2.000	100,00

Distribuição de frequências

Comparação de diferentes amostras

- É possível comparar a mesma variável para diferentes amostras;
- Para tanto não se pode utilizar a “Frequência” sim a **porcentagem!**
- Isso porque a porcentagem reduz as observações à uma mesma base, no caso, base 100...

Gráficos de Frequências

Para os próximos tópicos, vocês deverão fazer a distribuição de frequências para os dados de salários, com 5 classes;

Pode utilizar Excel ou qualquer outro programa que tenham facilidade. Se preferirem pode fazer esse trabalho à mão...

A distribuição de vocês deve conter a frequência relativa, a frequência acumulada, a porcentagem relativa e a porcentagem acumulada de cada classe.

Qual a média salarial? E qual o valor salarial que mais se repete nesta empresa? O que você pode inferir da distribuição de salários dessa empresa?

Gráficos de Frequências

O resultado será uma tabela parecida com essa:

Classe de salários	Frequência n_i	Frequência acumulada N_i	Porcentagem $100 f_i$	Porcentagem acumulada $100 F_i$
4,00 — 8,00	10	10	27,78	27,78
8,00 — 12,00	12	22	33,33	61,11
12,00 — 16,00	8	30	22,22	83,33
16,00 — 20,00	5	35	13,89	97,22
20,00 — 24,00	1	36	2,78	100,00
Total	36	—	100,00	—

Gráficos de Frequências

O gráfico de porcentagem acumulada fica como o gráfico ao lado

Mas o que ele representa?



Vamos ao Python...

Escolha um compilador de códigos Python à sua escolha, a saber que nesta disciplina, todas as aulas práticas deverão ser entregues no formato IPYNB (notebook jupyter)

Sugestões:

1. VS Code
2. Anaconda
3. Jupyter Notebook (web)
4. CoCalc (web)

Uma vez escolhido o seu compilador...

- Abra o arquivo “Aula Prática 1.ipynb”;
- As aulas com práticas serão realizadas diretamente no arquivo da aula prática;
- Ao finalizar a atividade (após ter realizado todas as atividades e respondido todas as questões disponíveis) você deverá salvar o arquivo com o seguinte esquema:

Aula Prática ?? – Seu nome.ipynb

- As atividades deverão ser entregues dentro do prazo estabelecido em aula