

# Noções de Estatística

## Aula 1

- Tipos de variáveis
- Distribuições de Frequências
- Representações Gráficas

### 1.1 Tipos de variáveis

- O que é uma variável?
- Por que é importante saber identificar o tipo de uma variável?

Na estatística, é muito importante ficar atento aos tipos de informações que poderão vir a ser úteis ao analisarmos algum fenômeno de interesse.<sup>1</sup>

Uma variável é qualquer característica que pode ser mensurada e/ou contada.

Exemplos de variáveis:

Número de Filhos

Sexo

Altura

Cor dos olhos

Tipo de veículo

Grau de Escolaridade

---

<sup>1</sup> Fenômeno de interesse: Qualquer fenômeno em que estejamos interessados em fazer previsões, ou obtermos informações sobre seu comportamento, ou compará-lo com um outro fenômeno, etc.

Ou seja, praticamente qualquer característica pode ser considerada uma variável.

A variável tem esse nome porque o valor pode variar entre as unidades amostrais de uma população ou porque podem sofrer alterações ao longo do tempo.

Observe que as variáveis listadas como exemplos podem representar características bem diferentes entre si.

Vamos observar mais atentamente cada uma delas.

- Número de filhos:

Com esta variável, esperamos observar algum valor numérico, geralmente um valor inteiro.

- Sexo:

Observe que esta característica é diferente da característica “número de filhos”. Nesta variável, esperamos observar um valor nominal, entre 2 possíveis valores Masculino ou Feminino.

- Altura:

Que tipo de valor esperamos observar nesta variável?

- Grau de Escolaridade:

Aqui esperamos observar novamente um valor nominal. Uma importante diferença entre essa característica e a característica Sexo é a possibilidade de haver uma certa hierarquia entre as respostas, ou seja, uma ordem.

Fundamental → Médio → Superior

Observe as demais características listadas e o que se espera obter como resposta nesses casos.

Com isso, observamos 4 grandes distinções entre as variáveis. Primeiramente observamos que a principal distinção entre duas variáveis é se ela é um número ou um nome. Às variáveis que são representadas por um número, emprega-se o nome de variáveis quantitativas. Às variáveis que são um nome, variáveis qualitativas.

Dentre as variáveis quantitativas, é importante observar outra grande distinção.

Variáveis discretas: Valores que resultam geralmente de uma contagem, portanto são geralmente números naturais. Exs: número de filhos, número de acidentes em uma rodovia, etc.

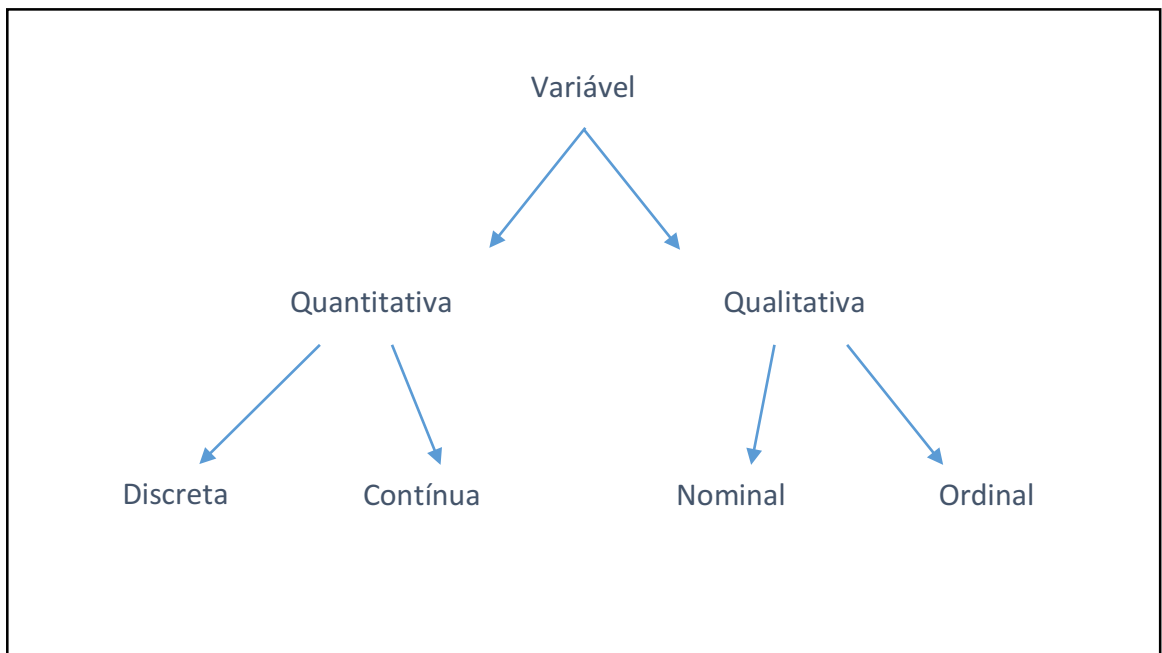
Variáveis contínuas: características mensuráveis que assumem valores em uma escala contínua (na reta real), para as quais valores fracionais fazem sentido. Usualmente devem ser medidas através de algum instrumento. Exemplos: peso (balança), altura (régua), tempo (relógio), pressão arterial, idade.

Dentre as variáveis qualitativas, é importante observar outra grande distinção.

Variáveis nominais: não existe ordenação dentre as categorias. Exemplos: sexo, cor dos olhos, fumante/não fumante, doente/sadio.

Variáveis ordinais: existe uma ordenação entre as categorias. Exemplos: escolaridade (1o, 2o, 3o graus), estágio da doença (inicial, intermediário, terminal), mês de observação (janeiro, fevereiro,..., dezembro).

Resumindo:



É muito importante saber fazer essa distinção entre as características observáveis em um estudo. O tipo da variável irá influenciar diretamente na sua análise. Diferentes tipos de variáveis são representados graficamente de maneiras diferentes.

---

Para cada tipo de variável existem técnicas apropriadas para resumir as informações.

Para finalizar, cabem duas observações:

Sobre as variáveis qualitativas:

Em algumas situações podem-se atribuir valores numéricos às várias qualidades ou atributos (ou, ainda,

classes) de uma variável qualitativa e depois proceder-se à análise como se ela fosse quantitativa, desde que o procedimento seja passível de interpretação.

Sobre as variáveis quantitativas:

Também é muito comum transformar variáveis quantitativas em qualitativas. Por exemplo, poderíamos ter uma variável quantitativa a qual gostaríamos de convertê-la em classes.

## 1.2 Distribuição de Frequências

Agora iremos iniciar o processo de análise estatística. A primeira parte de toda análise estatística tem início com uma análise exploratória dos dados. Essa etapa visa estudar, organizar, limpar e entender os dados do estudo.

Quando se estuda uma variável, o maior interesse do pesquisador é conhecer o comportamento dessa variável, analisando a ocorrência de suas possíveis realizações.

Vamos analisar uma tabela sobre uma pesquisa hipotética:

Tabela 1.1 Tabela de um estudo hipotético

Nº	Estado Civil	Grau de escolaridade	Nº de filhos	Salário (x sal min)	Idade	Região de procedência
1	Solteiro	E.f.	-	4,00	26	Interior
2	casado	E.M.	1	1,56	23	capital
3	Solteiro	E.S.	2	2,33	25	Interior
4	casado	E.f.	-	2,50	23	capital
5	Solteiro	E.M.	3	4,00	26	Interior
6	casado	E.S.	4	5,12	27	capital
7	Solteiro	E.f.	2	2,21	22	Interior
8	casado	E.M.	1	2,12	24	capital
9	Solteiro	E.S.	-	3,21	25	Interior
10	casado	E.f.	2	1,00	26	capital
11	Solteiro	E.M.	1	1,21	27	Interior
12	Solteiro	E.S.	3	1,00	32	capital
13	casado	E.f.	1	1,21	21	Interior
14	Solteiro	E.M.	2	1,21	31	capital
15	casado	E.S.	-	3,21	41	capital
16	Solteiro	E.f.	3	4,21	52	Interior
17	Solteiro	E.M.	2	3,21	53	capital
18	casado	E.S.	1	1,21	54	Interior
19	Solteiro	E.f.	-	1,86	34	capital
20	casado	E.M.	1	1,23	32	Interior
21	Solteiro	E.S.	2	1,32	42	capital
22	Solteiro	E.f.	-	15,22	26	Interior
23	casado	E.M.	1	1,23	43	capital
24	Solteiro	E.S.	-	1,43	43	Interior
25	casado	E.f.	2	1,32	35	capital
26	Solteiro	E.M.	1	1,54	45	capital
27	Solteiro	E.S.	2	1,43	34	Interior
28	casado	E.f.	1	1,43	54	capital
29	Solteiro	E.M.	2	1,53	34	Interior
30	casado	E.S.	1	1,43	44	capital
31	Solteiro	E.f.	1	1,65	43	Interior
32	casado	E.M.	1	1,56	23	capital
33	Casado	E.S.	1	1,43	54	Interior
34	Solteiro	E.f.	1	1,43	43	capital
35	casado	E.S.	1	1,32	34	Interior
36	solteiro	E.S.	1	1,76	54	capital

Em seguida, vamos procurar entender melhor o comportamento da variável Grau de Escolaridade. Para isso, utilizaremos um método muito comum de visualização gráfica que visa resumir os dados da tabela

para uma melhor compreensão a respeito do comportamento da variável de interesse. Esse método é a construção de uma tabela de distribuição de frequências.

Tabela 1.2 Distribuição de frequências da variável “grau de escolaridade”.

Grau de Escolaridade	Frequencia	Proporção	Porcentagem
E.F.	11	0,3055	30,55%
E.M.	12	0,3333	33,33%
E.S.	13	0,3611	36,11%
Total	36	1	100%

Observe que com essa tabela conseguimos visualizar um pouco melhor o comportamento dessa variável.

Vemos que, por exemplo, no nosso estudo há mais pessoas com grau de escolaridade Ensino Superior do que Ensino Fundamental. Mas também observamos que as porcentagens e contagens são muito próximas.

Como deveríamos construir uma tabela de frequências para a variável “Salário”?

Agrupar dados por faixas de salários.

Fazendo isso, estaremos resumindo os dados de uma variável contínua. Com isso, perde-se informação.

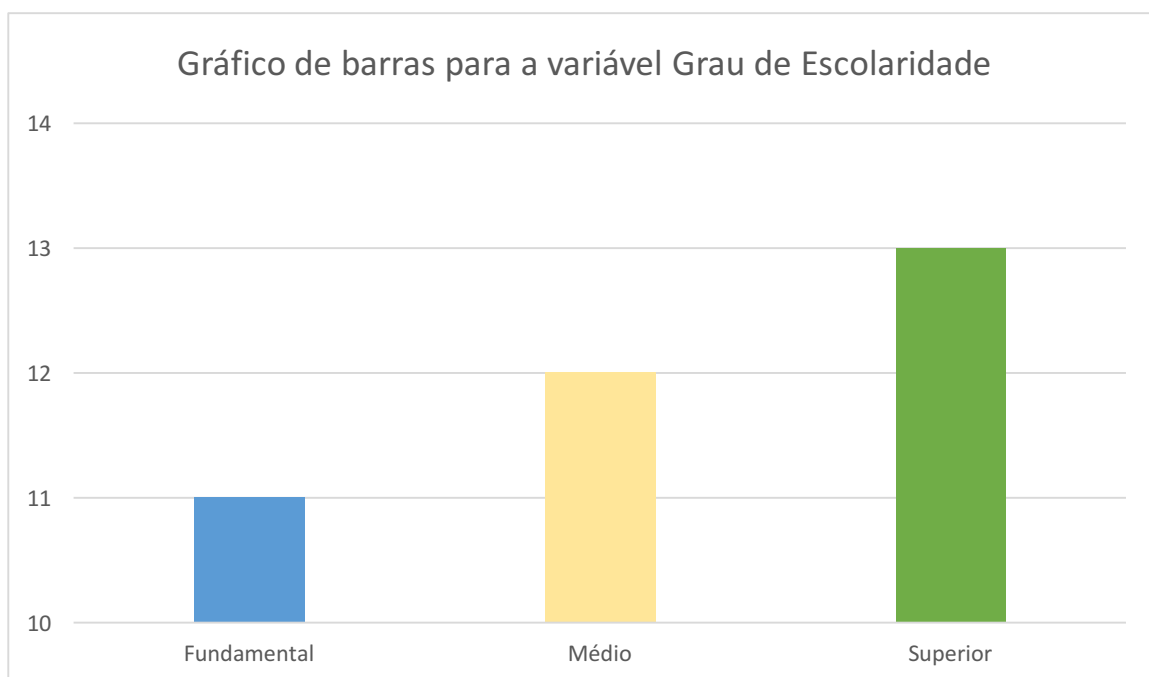
### 1.3 Gráficos

Observe como é difícil tirar conclusões interessantes apenas observando os dados da Tabela 1.1. Por isso, métodos de visualização gráfica foram desenvolvidos para facilitar a compreensão e investigação das variáveis envolvidas.

A seguir, será apresentada uma série de gráficos muito comuns e importantes para uma boa análise estatística preliminar.

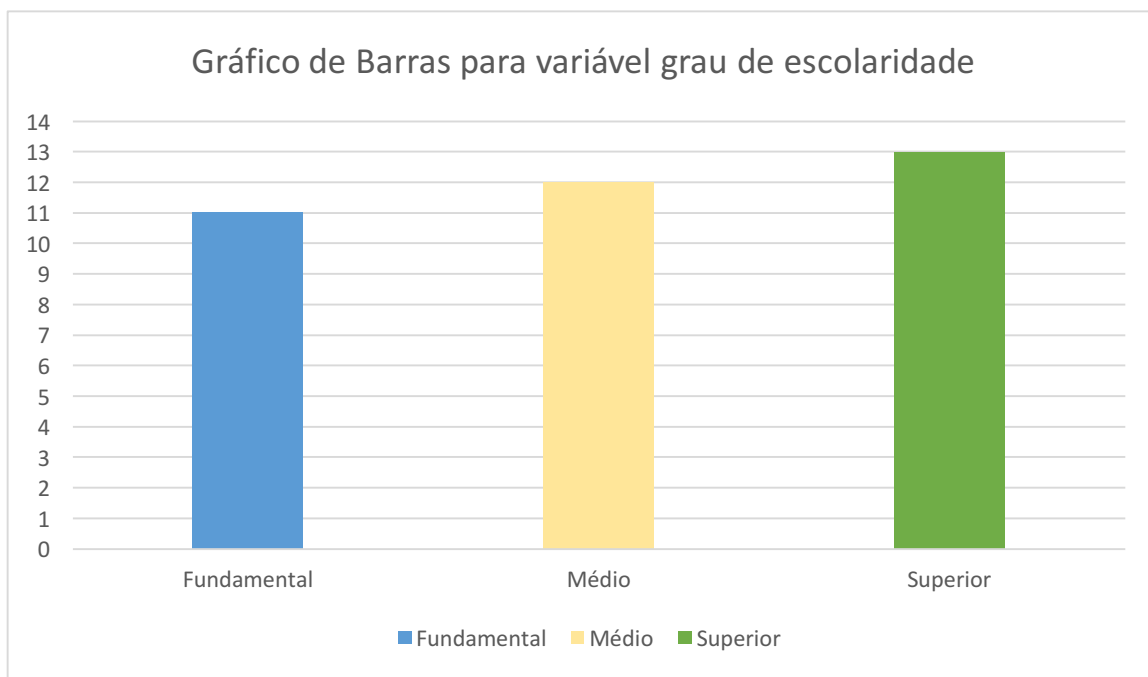
### 1.3.1 Gráficos para variáveis qualitativas

- Gráficos de barras



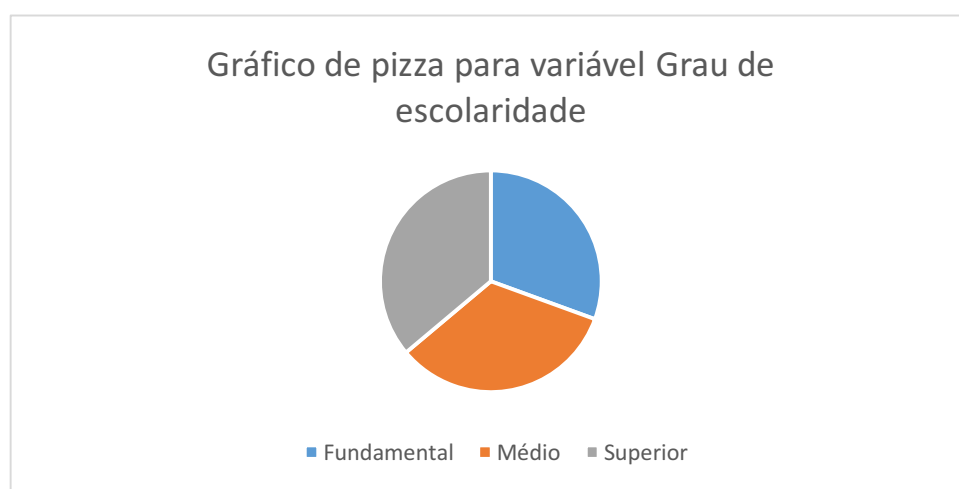
O gráfico de barras pode ser um excelente método para compararmos as diferentes frequências entre classes diferentes de uma mesma variável qualitativa. Mas fique atento à escala. Com essa escala, por exemplo, temos a impressão de que há muito mais indivíduos do ensino superior do que do fundamental. Observe outro gráfico de barras possível.





Observe a diferença com relação ao gráfico anterior. O gráfico anterior estava com o eixo-y iniciando a partir da contagem 10. Essa construção pode causar grande impacto na interpretação subjetiva de uma pessoa desatenta. Ao olhar para os dados de uma forma mais adequada, vemos que, de fato, não há muita diferença entre classes desta variável.

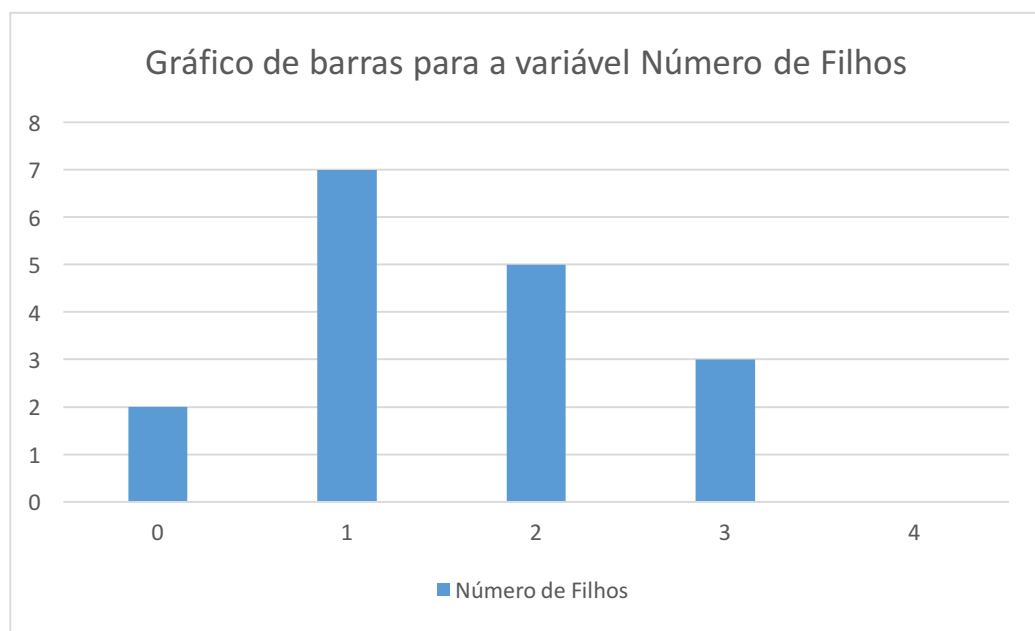
Alternativamente, pode-se utilizar o gráfico de pizza.



Observe como o gráfico de pizza reforça a conclusão que havíamos chegado com o segundo gráfico.

### 1.3.2 Gráficos para variáveis quantitativas

Para as variáveis discretas, podemos construir um gráfico de barras similar ao que construímos na seção 1.3.1.

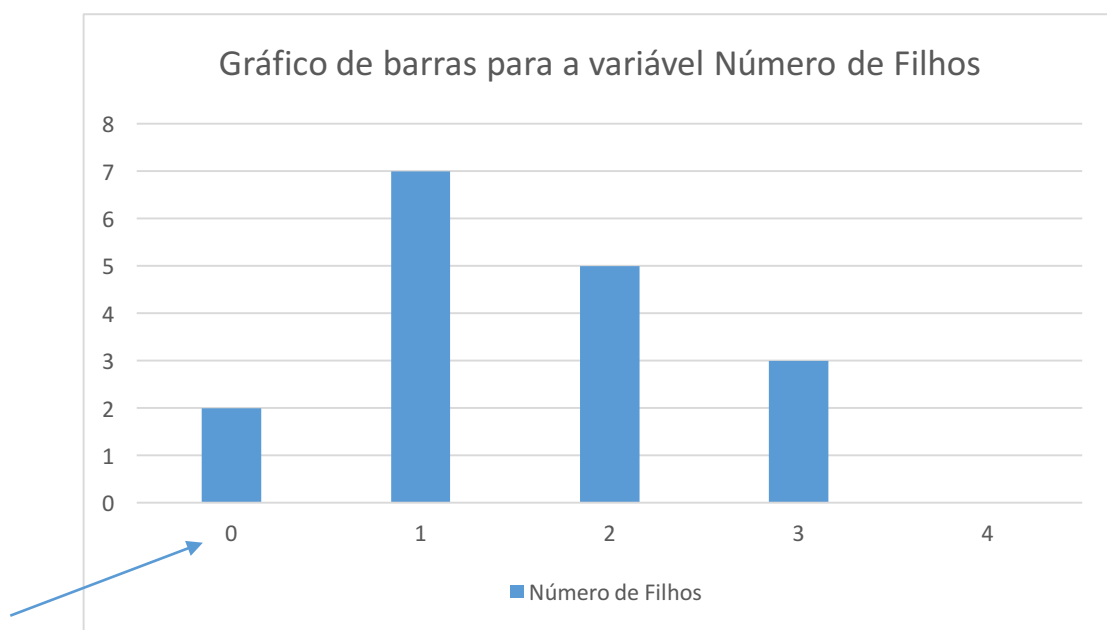


Para as variáveis contínuas, podemos criar classes, assim como na tabela de frequências. Porém, esse processo pode ser custoso pois perde-se informação ao se transformar variáveis quantitativas em classes. Por exemplo, observe a variável Salário na tabela 1.1. Poderemos tentar criar classes, mas elas podem nem sempre estar homogêneas.

Há uma alternativa para esse problema: O Histograma.

O Histograma é um gráfico de barras contíguas, com as bases proporcionais aos intervalos das classes e a área de cada retângulo proporcional à respectiva frequência. Ou seja, veja que no gráfico de barras comum não nos preocupamos muito com a largura das barras,

certo? E também, observe que para cada barra há apenas um número abaixo e centralizado.

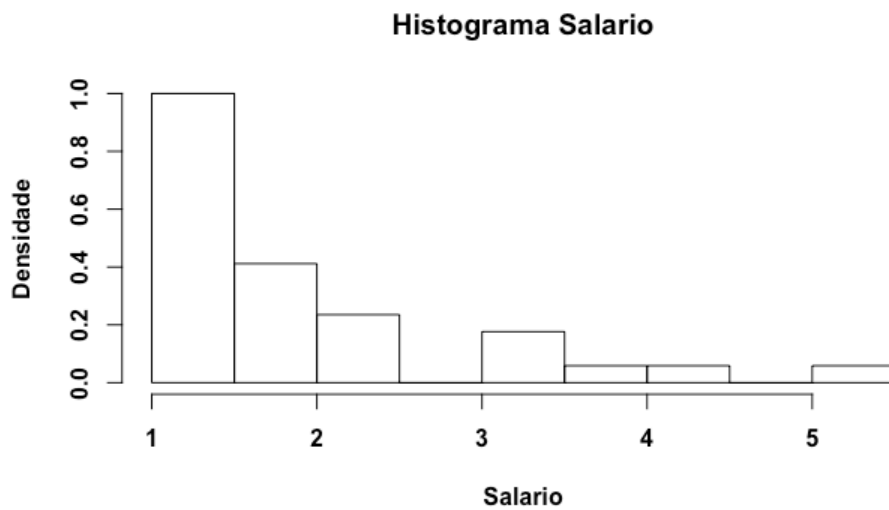


O Histograma é diferente.

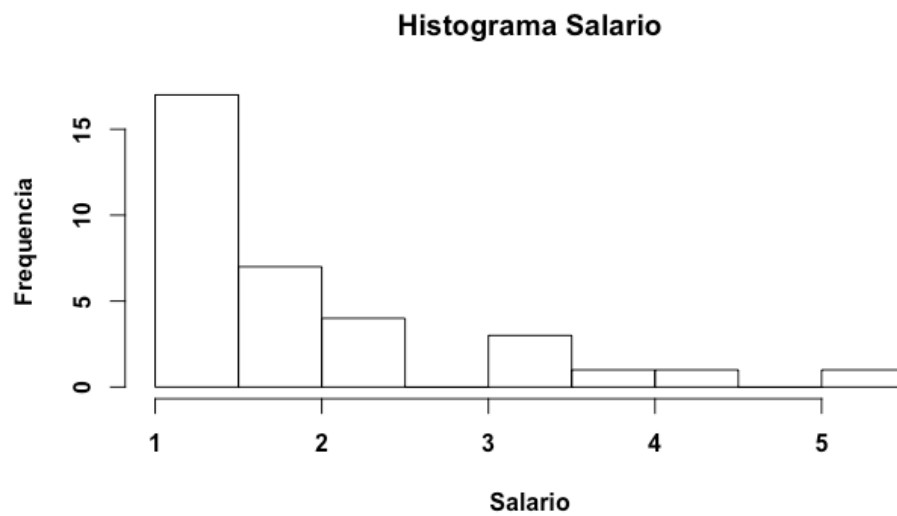
- Barras contíguas
- Bases proporcionais aos intervalos das classes
- Área de cada retângulo proporcional à respectiva frequência
- Pode-se utilizar a densidade de frequência no eixo-y (Nesse caso, a soma da área de todos os retângulos é igual a 1).

Exemplos de Histogramas para a variável Salário (lembrando que o valor numérico desta variável representa o número de salários mínimos):

Histograma com o eixo y representando a densidade de frequência, área total = 1.



Histograma com o eixo y representando a frequência absoluta.



Existem ainda muitos outros tipos de gráficos. Mas antes de avançarmos, precisamos estudar o próximo tópico: Medidas Resumo.