

Variáveis podem ser classificadas da seguinte forma:

1. **Variáveis Quantitativas:** são as características que podem ser medidas em uma escala quantitativa, ou seja, apresentam valores numéricos que fazem sentido. Podem ser contínuas ou discretas.
 1. **Variáveis discretas:** características mensuráveis que podem assumir apenas um número finito ou infinito contável de valores e, assim, somente fazem sentido valores inteiros. Geralmente são o resultado de contagens. Exemplos: número de filhos, número de bactérias por litro de leite, número de cigarros fumados por dia.
 2. **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.
2. **Variáveis Qualitativas (ou categóricas):** são as características que não possuem valores quantitativos, mas, ao contrário, são definidas por várias categorias, ou seja, representam uma classificação dos indivíduos. Podem ser nominais ou ordinais.
 1. **Variáveis nominais:** não existe ordenação dentre as categorias. Exemplos: sexo, cor dos olhos, fumante/não fumante, doente/sadio.
 2. **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).

Boxplot: Um resumo esquemático denominado boxplot é usado para descrever as características mais proeminentes de conjuntos de dados. Essas características incluem (1) centro, (2) dispersão, (3) a extensão e a natureza de qualquer desvio em relação à simetria e (4) a identificação de outliers, observações que normalmente estão distantes da maior parte dos dados

EXERCICIO 😊

Para resolver as questões apresentadas, vamos organizar os dados fornecidos:

2 1 2 4 0 1 3 2 0 5
3 3 1 3 2 4 7 0 2 3
1 0 4 2 1 3 1 2 5 0
2 3 2 1 3 2 2 8 4 5
1 3 1 6 0 3 3 3 6 1
2 3

a) Para determinar as frequências e frequências relativas dos valores observados de x (número de transdutores fora das especificações em um lote), podemos contar quantas vezes cada valor aparece na lista e então dividir pelo total de observações (que é 60 lotes).

| x (número de transdutores fora das especificações) | Frequência | Frequência Relativa |
|--|------------|-----------------------------|
| 0 | 10 | $10/60 = 1/6 \approx 0.167$ |
| 1 | 12 | $12/60 = 1/5 = 0.2$ |
| 2 | 10 | $10/60 = 1/6 \approx 0.167$ |
| 3 | 15 | $15/60 = 1/4 = 0.25$ |
| 4 | 6 | $6/60 = 1/10 = 0.1$ |
| 5 | 4 | $4/60 = 1/15 \approx 0.067$ |
| 6 | 2 | $2/60 = 1/30 \approx 0.033$ |
| 7 | 1 | $1/60 \approx 0.017$ |
| 8 | 0 | $0/60 = 0$ |

b) Para determinar as proporções solicitadas:

- Proporção de lotes na amostra que possuem no máximo cinco transdutores fora das especificações:

$$P(x \leq 5) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5) = 0.167 + 0.2 + 0.167 + 0.25 + 0.1 + 0.067 = 0.951$$

$$P(x \leq 5) = P(0) + P(1) + P(2) + P(3) + P(4) + P(5) = 0.167 + 0.2 + 0.167 + 0.25 + 0.1 + 0.067 = 0.951$$

- Proporção de lotes na amostra que possuem menos de cinco transdutores fora das especificações:

$$P(x < 5) = P(0) + P(1) + P(2) + P(3) + P(4) = 0.167 + 0.2 + 0.167 + 0.25 + 0.1 = 0.884$$

$$P(x < 5) = P(0) + P(1) + P(2) + P(3) + P(4) = 0.167 + 0.2 + 0.167 + 0.25 + 0.1 = 0.884$$

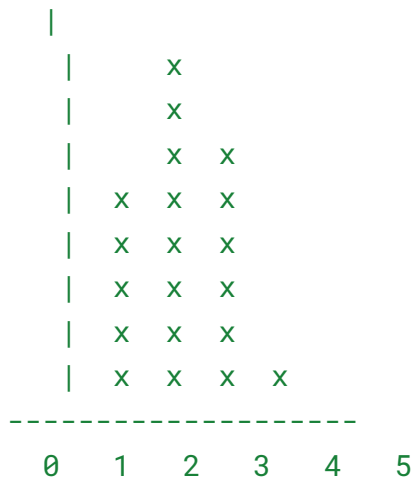
- Proporção de lotes na amostra que possuem no mínimo cinco transdutores fora das especificações:

$$P(x \geq 5) = 1 - P(x < 5) = 1 - 0.884 = 0.116$$

$$P(x \geq 5) = 1 - P(x < 5) = 1 - 0.884 = 0.116$$

c) Para desenhar um histograma dos dados, podemos utilizar as frequências relativas como alturas das barras. Vou criar uma representação gráfica:

lua



Esse histograma mostra a distribuição dos dados, onde cada barra representa a frequência relativa de cada valor de x. Nota-se que a maior frequência está em $x = 3$, seguida por $x = 1$.

Bom, a estatística é uma ciência de **análise de dados**. Aqui aprendemos a como coletar, organizar e analisar dados de forma que a gente possa tirar conclusões corretas a partir deles.

Pensa aí nas pesquisas eleitorais, nos estudos de como uma bactéria se propaga no corpo humano, entre várias outras coisas... Pra que a gente possa entender essas coisas, precisamos coletar dados e depois analisá-los!

Primeiro de tudo, quero que você entenda a diferença entre dois conceitos muito importantes, vem comigo!

População x Amostra

Vamos supor que você queira avaliar os níveis de violência da sua cidade, só que você só tem os dados relativos a roubos, furtos e essas coisas da sua rua.

Nesse caso:

População → todas as ruas da cidade.

Amostra → a sua rua.

Já deu pra você sentir o feeling da coisa, né?

Imagina aí um bolo bem grande e gostoso. Esse bolo inteiro seria a nossa população. Se a gente tirasse uma fatia dele, essa fatia seria a amostra. Pensa sempre nesse exemplo que você não esquece mais!

População é o conjunto dos elementos que você quer avaliar algumas propriedades. Como geralmente a população é gigante, ou até infinita, fica muito difícil avaliar ela como um todo. Surge então a amostra, que é uma pequena parte dessa população.

Outro ex: O datafolha e o Ibope fazem exatamente isso nas pesquisas de voto na eleição. Como é impossível pra eles saberem a opinião de toda a população brasileira, eles fazem uma pesquisa de intenção de votos com uma amostra da população e tiram conclusões sobre toda a população.

Sim, analisando os dados da amostra, você pode tirar conclusões sobre a população, e o nome disso é **inferência**! A gente vai começar a fazer isso o tempo todo a partir de agora.

Maass... pra isso, você tem que garantir que essa sua amostra é digna de confiança, que ela realmente representa bem a sua população.

Exemplo: suponha que você more na rua mais violenta da cidade; talvez a sua rua não seja uma boa amostra, concorda? Talvez ela não represente fielmente toda a sua cidade.

A gente precisa saber coletar bem uma amostra. Por isso é tão importante a gente estudar essa tal de estatística! Haha

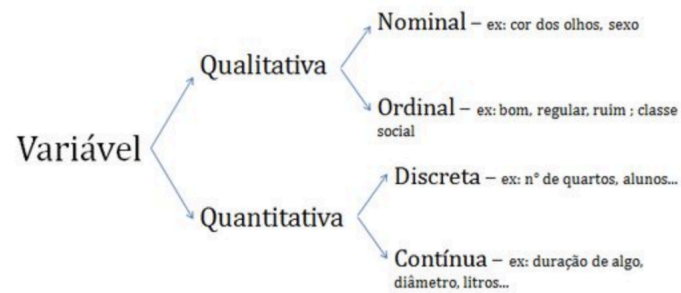
Observações:

- População não tem nada a ver com ser vivo não, pode ser população de geladeira, de cerveja, do que você tiver a fim.
- **Censo** é quando é feito um levantamento de toda a população, sem utilização de uma amostra. Por exemplo, a cada 10 anos, o IBGE faz um censo nacional, onde são pesquisados **TODOS** os domicílios brasileiros.

Tipos de Dados

Quando fazemos um levantamento de dados, sempre estamos analisando uma variável, concorda? Seja ela numérica (nº de quartos de um apartamento, nº de alunos, ...) ou não (avaliar algo em regular, bom, ruim, ...).

Podemos classificar essas variáveis da seguinte maneira:



FREQUÊNCIAS RELATIVA E ACUMULADA

Agora veremos outros tipos de frequências que existem. Utilizaremos elas para calcular as medidas de dispersão mais à frente.

A gente viu que frequência absoluta é a quantidade de ocorrências de um valor, né?! Além dessa, temos:

Frequência relativa (f_{ri}) será a frequência absoluta de um valor dividido pelo total de valores.

Voltando ao exemplo, se quisermos calcular a frequência relativa do $f_r(70)$, teremos:

$$f_r(70) = \frac{2}{20} = 0,1$$

Frequência acumulada (F_i) é o total acumulado (soma) de todas as frequências absolutas anteriores até a frequência absoluta do valor atual.

$$F_3 = f_1 + f_2 + f_3$$

No nosso exemplo, se quisermos calcular a frequência acumulada do valor 74 teremos

$$F(74) = f(70) + f(72) + f(74) = 2 + 1 + 1 = 4$$

Frequência acumulada relativa (F_{ri}) é você ir dividindo o valor da frequência acumulada pelo total de valores.

Voltando ao exemplo, se quisermos calcular a frequência acumulada relativa do 74, teremos:

$$F_r(74) = \frac{4}{20} = 0,2$$

Conselho: tomar muito cuidado pra não confundir esses f's todos haha.

| Peso | fi (frequência) | Fi (frequência acumulada) | fri (frequência relativa) | Fri (Frequência Acumulada Relativa) |
|------|--------------------|---------------------------------|---------------------------------|--|
| 70 | 2 | 2 | 0,1 | 0,1 |
| 72 | 1 | 3 | 0,05 | 0,15 |
| 74 | 1 | 4 | 0,05 | 0,2 |
| 75 | 2 | 6 | 0,1 | 0,3 |
| 76 | 1 | 7 | 0,05 | 0,35 |
| 77 | 3 | 10 | 0,15 | 0,5 |
| 78 | 1 | 11 | 0,05 | 0,55 |
| 80 | 1 | 12 | 0,05 | 0,6 |
| 82 | 1 | 13 | 0,05 | 0,65 |
| 83 | 1 | 14 | 0,05 | 0,7 |
| 84 | 2 | 16 | 0,1 | 0,8 |
| 85 | 2 | 18 | 0,1 | 0,9 |
| 86 | 1 | 19 | 0,05 | 0,95 |
| 88 | 1 | 20 | 0,05 | 1 |