

02

Qual é o tipo do seu dado?

A ideia deste curso é ensinar **Estatística Aplicada**, ou seja, queremos que você saia daqui sabendo muito mais do que é Média, Mediana ou Correlação, mas sim quando usar cada uma dessas coisas. Temos certeza que você já ouviu aquela frase: "Existem mentiras, grandes mentiras e estatística". E isso não é tão mentira, porque se você não souber escolher o teste certo que você vai aplicar, a função de média correta que você vai usar, você vai chegar em números que não explicam nada sobre o conjunto de dados que você tem.

E estatística hoje é fundamental. Nós temos um monte de informações, sejam elas na política, na economia, nos esportes ou mesmo na sua empresa. A sua empresa vende hoje para um monte de pessoas um monte de produtos diferentes; você pode extrair informações de lá. A estatística vai te ajudar a entender um pouco mais sobre os seus dados: o que você vende mais, o que você vende menos, vai te ajudar até a predizer que produto vai vender mais e quando vai vender mais.

Mas, obviamente, você vai ter que entender estatística a fundo. Aqui não ficaremos entrando naquele matematiquês de como cada teste estatístico funciona. Óbvio que haverá um pouco de matemática porque não tem como fugir disso, mas este curso não é para matemáticos, é um curso para quem quer usar estatística e aplicá-la no dia-a-dia.

A primeira coisa que temos que entender sobre estatística é entender sobre números, sobre os tipos de dados que estamos usando naquele momento. Imagine um formulário que você responde na rua sobre seu sexo:

Sexo:

- Masculino
- Feminino
- Não quero declarar

Este é um tipo de dado que chamamos de *Categórico*, pois temos categorias e um é diferente do outro e não possuem relação. Masculino não é melhor que Feminino e assim por diante.

Um outro tipo de dado comum é o que chamamos de dado *Ordinal*. Por exemplo, você acabou de fazer um curso e pedem para você avaliar o professor de 1 a 10:

Nota: ()1 ()2 ()3 ()4 ()5 ()6 ()7 ()8 ()9 ()10

Esse tipo de dado chamamos de *Ordinal* porque existe uma ordem: $1 < 2 < 3 < \dots < 10$, mas não conseguimos comparar a diferença do 1 para o 2. Existe o sentimento que de 1 para 2 é a mesma coisa que de 2 para 3, porém não conseguimos medir de maneira precisa. Esse intervalos podem variar de pessoa para pessoa. É diferente medir, por exemplo, uma temperatura: ao medir a temperatura de um ser humano e indicar 36 °C, saberemos que é preciso. De 36 para 37 a diferença é de exatamente um grau Celsius.

A temperatura, então, é um exemplo de dado *Intervalar*:

25°C -- 25,2°C -- 26°C

A diferença de um para o outro é mensurável e precisa.

Outro tipo de dado, menos comum, é o *Racional*. Ele é bem parecido com o *Intervalar*: ele é composto de números em ordem e podemos medir a diferença de um para o outro. Porém, nesse tipo de dado o 0 (zero) significa a ausência daquela coisa. Medindo a temperatura em Celsius, 0 °C significa que está frio, mas não a ausência de temperatura. Em graus Kelvin, 0 K significa a ausência de temperatura. Nos estudos em Física, faz sentido lidar com os dados *Racionais*, porém aqui em Estatística trabalharemos com os três primeiros tipos de dados: *Categórico*, *Ordinal* e *Intervalar*.

Saber isso é necessário porque dependendo do seu tipo de dado, você tem que escolher o método estatístico certo. Na próxima aula veremos se a Média Aritmética que aprendemos na escola faz sentido para qualquer tipo de dado. Conhecer o seu tipo de dado é fundamental.

Com isso claro, vamos continuar. Raramente em Estatística iremos analisar apenas um número. Nós a utilizamos porque temos um monte de dados e precisamos reduzi-los a um número que os traduza e que possamos entendê-los de maneira fácil.

Vamos começar, então, a agrupar dados.

Exemplo:

Temos os nomes dos alunos de uma escola e a quantidade de aulas que cada um assistiu:

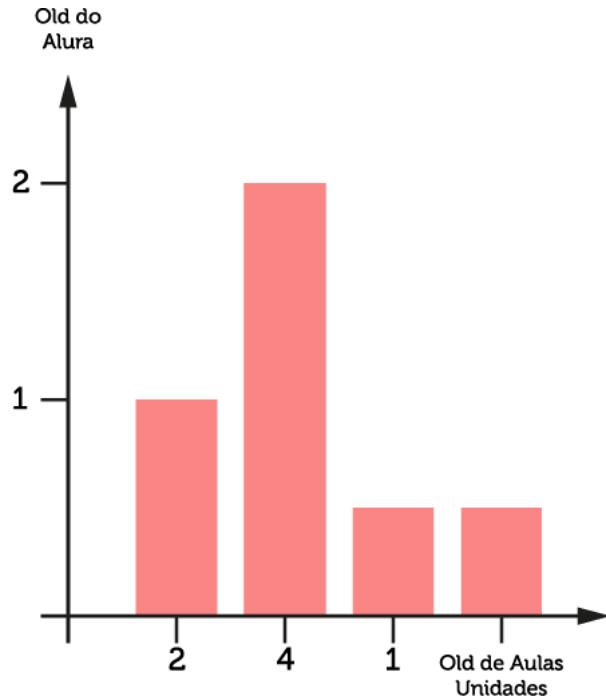
- Maurício: 2 aulas
- Natália: 4 aulas
- Felipe: 4 aulas
- João: 6 aulas
- José: 5 aulas

Agora queremos entender essas informações: quantas aulas os alunos assistem na escola? Para tal, utilizaremos um **Histograma**. O Histograma é um gráfico que mostra a quantidade de frequências e quantas vezes elas se repetem.

Quantidade de aulas assistidas	Quantidade de alunos
2	1
4	2
5	1
6	1

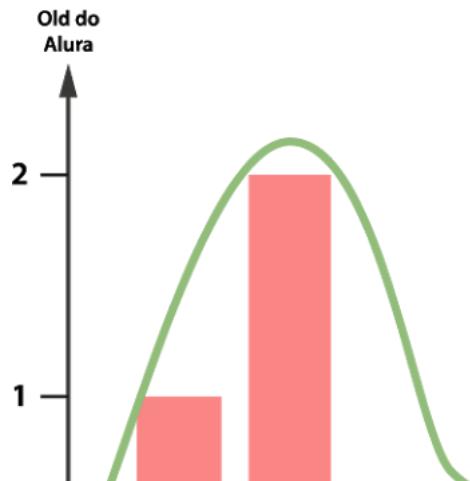
Perceba que temos a mesma informação só que disposta de outra maneira.

Porém, com essa tabela, conseguimos criar um gráfico:



Esse é o Histograma, o gráfico que mostra a frequência em que as coisas aconteceram em minha distribuição. Nesse exemplo utilizamos poucos dados, mas imaginemos uma distribuição maior. Deixaria bem claro o que está acontecendo, em relação às aulas assistidas, na escola.

Se traçarmos uma linha por esse gráfico:



Essa curva formada é importantíssima e é o que chamamos de *Curva Normal*. É essa curva que esperamos numa distribuição comum. Se, por exemplo, desenharmos o Histograma da altura de um homem brasileiro, veremos que poucos estarão nas faixas menores e maiores e muitos nas faixas centrais. Muita gente estará na média e pouca nos dois extremos. Entender se sua distribuição está dentro ou não dessa curva também é de extrema importância para escolhermos o teste estatístico ideal. Falaremos mais para frente sobre ela.

Revendo:

- antes de estudarmos a Média, ou o Teste de Correlação, ou Teste de Hipótese (veremos tudo isso nas próximas aulas), devemos entender os tipos de dados: *Categóricos, Ordinais e Intervalares*.
- Depois pensamos no formato da distribuição. Uma maneira fácil é desenhando um *Histograma*, que nos mostra a frequência da distribuição. Os testes a serem aplicados mudam se ela for Normal ou não.

Com essas informações na cabeça, você está pronto para começar a entender mais sobre estatística. Até a próxima aula!