



# Introdução à Estatística

## Introdução

Neste módulo, começaremos nosso estudo da Estatística e suas aplicações. Vamos ver o que é Estatística e como se aplica no mercado. Além disso, iremos dividir a Estatística em duas partes, a indutiva e a dedutiva, e vamos falar sobre cada uma delas. Por fim, veremos como conduzir um estudo usando o método estatístico, explicando cada uma de suas fases.

## Objetivos da aula

- Dar um panorama geral do que é Estatística e de como ela é utilizada no mercado
- Entender a diferença entre Estatística indutiva e dedutiva
- Explicar o método estatístico e suas fases, além dos benefícios de sua aplicação

## Resumo

Podemos definir Estatística como a ciência de usar informação obtida através da coleta, organização e estudo de dados. É um braço



matemática que nos permite tirar insights sobre quaisquer tipos de dados que tivermos à nossa disposição. 


Aliás, essa é a palavra chave, **dados**. Como a famosa frase diz, "Sem dados, você é só mais uma pessoa com opinião". Hoje, tomar decisões baseadas em dados deixou de ser um diferencial competitivo e se tornou um must have para qualquer empresa que quer ser bem sucedida. Nesse contexto, o uso da Estatística vem se tornando cada vez mais relevante, sendo objeto de estudo até mesmo entre as pessoas executivas, visando tomadas de decisões mais conscientes e informadas.

Existem inúmeras aplicações da Estatística tanto em nosso dia a dia quanto no mercado. Quando falamos, por exemplo, da eficácia de vacinas, estamos usando estatística para prever se uma vacina vai ser suficiente para gerar imunidade de rebanho ou não. Quando vemos as prévias eleitorais e falamos sobre margem de erro, é estatística pura sendo aplicada a um assunto tão importante para a sociedade.

Já no mercado, a Estatística tem sido usada há décadas não só para tomada de decisão, mas também como modelos de previsão e estudos de comportamento. Por exemplo, quando pensamos em análise de mercado, logo nos vem à mente indicadores de métricas importantes e dashboards no Tableau, Power BI e por aí vai.

Fonte: <https://i.ytimg.com/vi/r2WvjRjTPrw/maxresdefault.jpg> (acesso em 08/12/2022)


Esses indicadores, em geral, são medidas estatísticas que nos ajudam a sumarizar a informação; entre elas, a mais famosa de todas, a média! Mas não para por aí, vemos Estatística aplicada a detecção de anomalias, onde um grande caso de uso é em gastos de cartão de crédito.

Pense que você possui um cartão de crédito que usa regularmente e, em alguma ocasião, os dados do seu cartão foram roubados. Quando  alguém tentar utilizar o seu cartão para, por exemplo, uma compra online, essa compra será fraudulenta. A operadora do seu cartão de crédito vai analisar essa compra e, se ela estiver fora do seu padrão de consumo, ela pode escolher bloquear essa compra, prevenindo uma fraude. Esse padrão de consumo, bem como a decisão de classificar essa compra como fraudulenta pode ser feita usando um método estatístico de detecção de anomalias.

Uma outra importante aplicação é o Teste de Hipótese. Com ele, podemos conduzir experimentos controlados para testar ideias que nossas equipes desenvolvam e analisar quais situações são melhores. Por exemplo, imagine que você trabalha em uma empresa de varejo e está testando um novo fluxo no checkout do carrinho do site de vendas. Um teste de hipótese, feito após um experimento bem conduzido, vai nos ajudar a escolher qual fluxo trará, por exemplo, uma maior completude de vendas no checkout. Todas essas aplicações da Estatística podem ser classificadas dentro de duas divisões da disciplina, a Estatística dedutiva e a indutiva.

A Estatística dedutiva também é conhecida como Estatística descritiva e seu objetivo é obter informações e insights sobre o que já foi observado. Por exemplo, quando queremos conduzir um estudo sobre o padrão de consumo das pessoas que utilizam um serviço de streaming de vídeo, entendendo a variedade de títulos, média de tempo visto e até mesmo informações demográficas, estamos falando de Estatística descritiva.

A Estatística indutiva também é conhecida como Estatística Inferencial e seu objetivo é obter estimativas para o que não foi observado, mas utilizando o que já foi observado. Utilizando novamente o exemplo do serviço de streaming de vídeo, uma aplicação da Estatística Inferencial


seria tentar prever quais conteúdos farão sucesso utilizando os dados de consumo já observado, ou mesmo tentar classificar potenciais ,  joas que podem cancelar o serviço a qualquer momento e tentar fazer alguma ação para que permaneçam com suas assinaturas.

Tanto utilizando Estatística inferencial quanto descritiva, pode ser difícil de encontrar soluções, estimativas e insights utilizando os dados que temos. Por isso, se faz necessário um processo para facilitar a análise estatística dos dados que temos à nossa disposição. Na prática, vamos utilizar o Método Estatístico e suas 7 fases para descomplicar a análise de nossos dados.

A primeira fase do método estatístico é a definição do problema. Essa é a fase mais importante de qualquer estudo analítico, pois vai direcionar completamente a análise que será feita. Em geral, stakeholders fazem pedidos e cabe à pessoa que conduzirá a análise entender qual o problema a ser definido por trás de cada pedido, fazendo perguntas e contextualizando com informações do negócio ou da pesquisa.

A segunda fase é a do planejamento. Nessa fase, traçamos um plano para resolver o problema que temos. Aqui, definimos quais dados vamos precisar, de onde vamos extraí-los, quais análises vamos fazer em cima desses dados e por quanto tempo vamos analisá-los. Essa última parte é importante pois podemos passar quanto tempo quisermos analisando nossos dados. Em geral, quanto mais tempo, mais profunda e detalhista é a análise, mas precisamos sempre pensar no timing de entrega do resultado.

A terceira fase é a da coleta dos dados. Aqui, vamos atrás dos dados que definimos que precisamos na fase anterior. Em qual base de dados eles estão, em qual formato, qual acesso precisamos e qual a natureza desses dados? São todas perguntas que precisamos responder nessa etapa.

A quarta fase é a da crítica dos dados. Essa parte é muito importante pois é quando investigamos a qualidade dos dados que temos.  Nessa disposição. Dados ruins levam a interpretações erradas. Por isso, é de suma importância que revisemos os dados que temos e seja feita uma checagem dos dados, entendendo se eles estão se comportando de acordo com o esperado. Por exemplo, se temos um dado de tempo gasto numa página, ele nunca pode ser negativo.

A quinta fase é a da apuração dos dados. Nessa fase, vamos revisar os dados que temos, com uma boa qualidade já assegurada, e começamos a fazer as análises que já planejamos fazer. Também é nessa fase que olhamos novamente para as perguntas que queremos responder e entendemos se falta algum dado ou alguma informação ou análise para cumprir esse objetivo.

A sexta fase é a de apresentação dos dados. Aqui, é importante entender o público a quem será apresentado essa análise: para um público não tão técnico, faz sentido usar termos de negócio e em geral, com mais gráfico e mais mastigado; para um público mais técnico, talvez já faça sentido levar análises mais cruas, explicando até mesmo como a análise foi feita. Além disso, é relevante entender a periodicidade da apresentação dos dados. Se forem analisados diariamente, ou semanalmente, por exemplo, vale a pena medir um esforço de uso de uma solução de dashboard mas, se for um estudo pontual, uma apresentação talvez faça mais sentido.

Por fim, a sétima e última fase é a da análise e interpretação dos dados. Nessa fase, conectamos o resultado das análises (gráficos, médias, estimativas, etc) com o problema de negócio que queríamos resolver em primeiro lugar. Um número ou um gráfico por si só não responde nenhuma pergunta, eles precisam estar ligados a um contexto de negócio e você é a melhor pessoa para fazer isso!

**Como aplicar na prática o que aprendeu**

Mais do que fazer conta, a parte mais importante da estatística é a análise de dados, momento em que você vai colocar todo o seu conhecimento sobre o assunto em prática. Para essa prática, analisar o dashboard disponibilizado aqui no próprio material: <https://i.ytimg.com/vi/r2WvjRjTPrw/maxresdefault.jpg>

Uma dica é fazer uma interpretação de cada quadro, extraindo os dados mais relevantes e interpretando os que têm menor ocorrência, também.

## Conteúdo bônus

### Tópicos avançados

Nesse artigo, você consegue encontrar mais detalhes da importância da Estatística na Ciência de Dados, até mesmo para Machine Learning.

Caro estudante, você consegue acessar os códigos utilizados na disciplina no link a seguir: <https://github.com/FaculdadeDescomplica/Statistics-for-Data-Science>

### Referência Bibliográfica

Método estatístico: o que é como se aplica à cadeia de suprimentos. Disponível em: <<https://neogrid.com/br/blog/metodo-estatistico-como-se-aplica-cadeia-de-suprimentos> > (Acesso em 01/12/2022)

**Ir para questão**

