

RF

Receita Federal

Estatística

SUMÁRIO

ESTATÍSTICA	5
■ ESTATÍSTICA.....	5
ESTATÍSTICA DESCRITIVA	5
■ MÉDIA.....	8
MODA.....	10
MEDIANA.....	10
■ DESVIO PADRÃO	11
■ PROBABILIDADE	12
DISTRIBUIÇÕES DE PROBABILIDADE	13
■ INFERÊNCIA.....	15
ESTIMAÇÃO PONTUAL	15
INTERVALAR.....	17
TESTES DE HIPÓTESES.....	20
■ PREDIÇÃO	25
ABORDAGENS.....	25
SÉRIES TEMPORAIS	25
REGRESSÃO LINEAR SIMPLES E MÚLTIPLA.....	27
REGRESSÃO LOGÍSTICA	36

ESTATÍSTICA

ESTATÍSTICA

ESTATÍSTICA DESCRITIVA

A **estatística** é a parte da matemática que se dedica à análise, apresentação e interpretação de dados coletados. Esses dados são coletados dentro de uma **população**, que é o conjunto total dos elementos a serem estudados (podem ser pessoas, objetos etc.). Dessa população, podemos coletar os dados de duas maneiras:

- **Censo:** quando são coletados os dados de toda a população; e
- **Mostra:** é um subconjunto da população, da qual são coletados dados para, posteriormente, fazer uma inferência sobre a população (inferência estatística).

Ex.: na eleição, todas as pessoas aptas a votar são a população. Quando uma empresa é contratada para fazer uma pesquisa de intenção de voto, eles selecionam uma amostra dessa população, fazem as perguntas predeterminadas pela pesquisa, e com os dados das respostas fazem uma inferência de como a população toda irá votar.

Um dos ramos da estatística é a **estatística descritiva**, onde estudaremos 4 tipos de medidas descritivas:

- As **medidas de tendência central** que são medidas que indicam a posição dos dados, como média, mediana, moda e quartis (também chamadas de **medidas de posição**);
- As **medidas de dispersão** que medem o grau de variabilidade dos elementos de um conjunto, como desvio-padrão, variância, amplitude;
- A **assimetria** da curva; e
- O achatamento da curva, chamado de **curtose**.

Importante!

Essas duas últimas (assimetria e curtose) também são conhecidas como **medidas de distribuição**.

Os dados de uma amostra podem ser **qualitativos**, que são aqueles dados não numéricos como sexo, nacionalidade, avaliação nominal (bom, regular, ruim) etc., ou **quantitativos**, que são dados expressos em números, que podem ser objeto de contagens, medições como altura, peso etc.

Cuidado, não necessariamente dados expressos em números serão quantitativos, eles podem também ser qualitativos, como RG, CPF, CNPJ, CEP, CNAE (classificação nacional de atividades econômicas), geralmente esse tipo de código ou classificação é feito em números. Ex: queremos saber a quantidade de empresas que atuam em cada setor, podemos usar a CNAE, que é um número, para separar a quantidade de mercados, farmácias, postos de combustíveis etc.

Os dados qualitativos podem ser:

- **Ordinais:** são aqueles que podem ser ordenados, como mês, nível de escolaridade, tamanho de roupa (P, M, G) entre outros. Ex: o nível de escolaridade pode ser dividido em ensino fundamental, ensino médio, ensino superior, pós graduação. Por mais que seja um dado qualitativo, a gente consegue colocar isso em ordem, pois sabemos que primeiro vem o ensino fundamental, depois o ensino médio e assim por diante. Da mesma forma o mês, sabemos que primeiro vem janeiro, depois fevereiro até chegar em dezembro;
- **Nominais:** que são aqueles que não podem ser ordenados, como sexo, estado civil entre outros. Ex: podemos dividir os estados civis em casado, união estável, solteiro, viúvo... claramente não temos uma ordem entre essas opções.

Os dados quantitativos podem ser:

- **Discretos:** são aqueles dados que possuem um conjunto finito de valores, como a quantidade de acertos em uma prova de múltipla escolha, a quantidade será apenas números inteiros, 0, 1, 2, 3 e assim por diante, ou
- **Contínuos:** que são aqueles que possuem uma escala contínua de valor como tempo, comprimento etc. Ex: vamos considerar a variável altura. Entre os dados 1,70m e 1,71m existe uma infinidade de números.

Normas de Apresentação Tabular

- Modelo de uma Tabela

Já mostramos algumas tabelas ao longo do material, mas afinal, para que serve, e como montar uma tabela?

Uma tabela deve ser composta por diversas linhas e colunas, sendo que devemos ter um título (normalmente na primeira linha da tabela), e vários dados organizados nas linhas e colunas seguintes.

Geralmente, na primeira linha, depois do título, teremos as classes que serão retratadas nas linhas, ex.: Estados, Siglas, População, Área, etc. Nas linhas seguintes teremos os dados da tabela, onde teremos em uma mesma linha o Estado, relacionando sua sigla, sua população, sua área etc. Vejamos o exemplo dessa tabela citada.

INFORMAÇÕES DOS ESTADOS DA REGIÃO SUDESTE			
ESTADO	SIGLA	POPULAÇÃO	ÁREA (KM ²)
Minas Gerais	MG	21.292.666	586.522
Espírito Santo	ES	4.064.052	46.095

INFORMAÇÕES DOS ESTADOS DA REGIÃO SUDESTE			
ESTADO	SIGLA	POPULAÇÃO	ÁREA (KM ²)
Rio de Janeiro	RJ	17.366.189	43.780
São Paulo	SP	45.919.049	248.222

Analisando a tabela acima, podemos concluir que Minas Gerais (MG) é o maior estado da Região Sudeste, pois tem a maior área, mas que o estado de São Paulo é o mais populoso, por ter uma população maior que os outros.

O importante é olhar uma tabela e entender quais dados podemos extrair com o que está apresentado nela.

As tabelas mais utilizadas na estatística são as tabelas de frequência, conforme apresentamos no item de média aritmética para dados agrupados.

- Tipos de Séries Estatísticas

As séries estatísticas são as diversas maneiras de apresentar os dados desejados em forma de tabela, o objetivo das séries estatísticas é organizar os dados observados e mostrá-los de maneira organizada, facilitando sua compreensão.

Temos vários tipos séries estatísticas, mas vamos destacar algumas mais importantes:

- **Séries Temporais:** é um conjunto de observações de uma variável ao longo do tempo, ou seja, uma sequência de dados numéricos em ordem sucessiva. Nesse tipo de série o que varia é o tempo, mas o fato e o local de observação são fixos;
- **Séries Geográficas:** é um conjunto de observações de uma variável em diferentes locais. Nesse tipo de série o que varia é o local (região) da observação, mas o tempo e o fato observado são fixos. Ex.:

POPULAÇÃO DOS ESTADOS DA REGIÃO SUDESTE	
ESTADO	POPULAÇÃO
Minas Gerais	21.292.666
Espírito Santo	4.064.052
Rio de Janeiro	17.366.189
São Paulo	45.919.049

- **Séries Específicas:** é um conjunto de observações de uma variável com diferentes categorias (espécies). Nesse tipo de variável o que varia são as categorias observadas, mas o tempo e o local são fixos. Ex.: queremos analisar a quantidade de animais diferentes que habitam uma certa região de proteção florestal. Nesse caso a tabela será classificada pelas espécies observadas na região de interesse em um mesmo intervalo de tempo.

ESPÉCIES QUE HABITAM A REGIÃO EM 2020	
ESPÉCIE	QUANTIDADE OBSERVADA
Onça	45