

Variância e Covariância

Mentora: Fernanda Fiel

Para aprofundar um pouco mais nossos conhecimentos em estatística, a expert [Fernanda Peres](#) vai explicar nesta aula alguns conceitos importantes relacionados ao comportamento dos dados, em particular como mensurar a variação de um tipo de dado e como podemos mensurar a associação entre dois dados distintos.

Chegou o momento de falarmos sobre Variância e Covariância.

Antes de começar, entenda quais são os objetivos de aprendizagem desta aula.

Objetivos de Aprendizagem

- Compreender conceitualmente o que é variância e como calculá-la;
- Entender o que é desvio padrão e como se relaciona com a variância;
- Compreender conceitualmente o que é covariância e como calculá-la;
- Entender a correlação como forma de mensurar a associação entre duas variáveis;
- Compreender a diferença entre correlação e causalidade.

Reserve um espaço de tempo para acompanhar a aula sem interrupções, sente-se em um lugar confortável e tenha papel ou editor de texto à mão para fazer anotações.

O que é Variância?

Neste capítulo, Fernanda fala sobre como interpretamos e calculamos a variância de um dado. Essa medida, que numericamente nos explica a dispersão dos dados, é uma das mais importantes estatísticas descritivas.

Resumo

- A variância é uma medida que quantifica o quanto os dados estão distantes da média. Quanto maior a variância, mais dispersos os dados estão.
- A variância amostral e a variância populacional são calculadas de formas diferentes, é preciso ter essa atenção na hora de executar esse cálculo manualmente;
- A variância é uma medida quadrática, isto é, uma medida que tem seu valor elevado ao quadrado. Por esse motivo, pode ser de difícil interpretação.

Como alternativa, temos o desvio padrão, que representa uma medida de variação na mesma escala do dado original.

O que é Covariância?

Assim como a variância, a covariância é uma medida extremamente importante quando queremos descrever o comportamento dos dados. Vamos ver nesse capítulo como a covariância nos ajuda a compreender o comportamento conjunto de duas variáveis, de modo que ela nos permite ver como elas estão associadas.

Resumo

- A covariância é uma forma de quantificar uma associação entre duas variáveis numéricas;
- Assim como a variância, a covariância pode ser amostral ou populacional, e as formas como se calcula cada uma delas é diferente;
- A covariância pode indicar se a associação entre duas variáveis é positiva ou negativa. No caso em que a covariância é positiva, quando uma das variáveis tem seu valor crescente a outra também terá (os valores crescem juntos). No caso em que a covariância é negativa, quando uma das variáveis tem seu valor crescente, a outra terá seu valor decrescente - e vice versa.
- A covariância mede se a associação entre duas variáveis é positiva ou negativa, mas ela **não** mede a magnitude dessa associação.

O COEFICIENTE DE CORRELAÇÃO DE PEARSON (R)

Como consequência direta das medidas de covariância e desvio padrão, o coeficiente de correlação nos permite mensurar o quanto duas variáveis estão numericamente associadas. Neste capítulo, vamos ver como calculamos essa correlação e como interpretamos essa informação.

Resumo

- Na covariância não é possível interpretar a magnitude com que dois dados se relacionam, mas derivada da covariância temos o coeficiente de correlação de Pearson que permite interpretar o quanto duas variáveis estão associadas numericamente.

- O coeficiente de correlação pode ser calculado a partir das medidas de covariância e desvio padrão das variáveis de interesse e sempre vai resultar em um número entre -1 e 1, que serve como escala para mensurar essa correlação;
- A interpretação da correlação se dá em duas partes:
 - 1) o sinal do valor da correlação, assim como no caso da covariância, indica se a associação entre as variáveis é positiva ou negativa e
 - 2) quanto mais próximo de -1 ou 1, mais forte é a correlação entre as variáveis. Uma correlação próxima de 0 indica uma correlação fraca;
- Correlação não implica em causalidade, duas informações estarem correlacionadas não implica que uma está causando a outra. Correlação é uma medida de associação **estritamente numérica**, isto implica que não é possível utilizar esse tipo de correlação para dados categóricos.

Resumo do Bloco

Para ajudar a consolidar as informações mais relevantes da aula, a Fernanda preparou uma lista de "takeaways", informações e dicas que você precisa levar com você:

Lembre-se

- **Variância** é uma forma de medir a dispersão dos dados;
- **Covariância** é uma forma de medir a associação entre duas variáveis;
- **Coeficiente de correlação** é uma medida padronizada de associação entre duas variáveis

Você concluiu esta aula

Que bom que você chegou até aqui.

Este é só o início do seu desenvolvimento de estatísticas descritivas.

Revise os aprendizados desta aula e avalie se você já tem segurança para colocar em prática o que aprendeu.

Checklist

Você deve sair desta aula entendendo:

- O que é variância e desvio padrão e o que essas informações explicam de uma variável;
- O que é a covariância e o que podemos concluir a partir dessa informação entre dois dados;
- Como utilizamos o coeficiente de correlação para mensurar como duas variáveis estão associadas;