

Desvio padrão

Transcrição

[0:00] Pessoal, vamos falar agora do desvio padrão, a medida de dispersão mais famosa, todo mundo deve ter escutado pelo menos falar do desvio padrão. Ele é, justamente, a raiz quadrada da variância, que vimos no vídeo anterior, você viu que a variância, quando a gente calcula ela, é uma média dos desvios ao quadrado, isso faz com que a unidade de medida dessa estatística descritiva seja elevada ao quadrado também, como dei o exemplo na aula anterior, a gente estava falando da altura, está em metros, a variância da altura vai estar em metros ao quadrado, ou seja, unidade de medida diferente, isso dificulta a interpretação.

[0:36] No caso do desvio padrão, não, a gente pega esse valor da variância, passa raiz quadrada nele e resolve o problema da unidade de medida, ele era em metros ao quadrado, volta a ser em metros, e facilita nossa interpretação, vamos ter uma medida de dispersão média dos dados, na mesma unidade de medida do próprio dado que gerou, ok?

[0:59] Vamos simplesmente calcular aqui uma variância e fazer comparações com nosso dataframe de exemplo pra entender melhor.

[1:14] Pra calcular, lembra que a gente importou no começo do curso o numpy, esse sujeito, deu apelidinho pra ele de NP.

[1:22] Numpy tem uma função matemática pra extrair raiz quadrada, sqrt.

[1:32] E vou passar aqui a variância que calculamos na aula anterior. Variância, calculamos a variância das notas de fulano, então vai ser um desvio padrão das notas de fulano, 2.3 o desvio padrão.

[1:51] Outra forma de obter isso, vamos chamar aqui, desvio padrão, vai ser igual a notas fulano, lembrando de pegar aqui as notas, somente, de Fulano, e passar aqui o STD, que é a função para calcular desvio padrão, agora não vou me esquecer de mostrar pra vocês o resultado, lembrando que do mesmo jeito da variância, os dois valores são iguais, e aqui estamos calculando o desvio padrão amostral.

[2:33] Da mesma forma que calculamos a variância amostral, lembra que fizemos na mão, e dividimos por n-1, o fator de correção que falamos antes, tínhamos esse valor, usando o VAR do Pandas, tivemos o mesmo valor, o Pandas está calculando variância amostral.

[2:50] Eu passando raiz quadrada nessa variância, vou estar calculando desvio padrão amostral, desse jeito, esse miolo é justamente a variância populacional, e esse miolo é variância amostral que vimos na aula anterior.

[3:06] Então é isso, aqui temos a forma de calcular usando STD, e agora, eu queria voltar no nosso dataframe pra termos uma ideia dessa brincadeira, porque depois desse vídeo, você já vai colocar a mão na massa e fazer análise descritiva do conjunto de dados, você está achando que não estou falando do conjunto de dados, porque estou tentando mostrar pra você como funciona pra você colocar a mão na massa e tirar suas conclusões do conjunto de dados.

[3:33] Mostrei aqui o DF, a gente pode mostrar de novo, .min, a gente vê as médias e vê que elas são, pro caso de fulano e ciclano, são idênticas, basicamente nisso que estou querendo basear meu exemplo, em analisar fulano e ciclano. Aqui no caso, a mediana dos caras é igual, vamos fazer a moda.

[4:03] .mode. Está lá, olha que coisa interessante, você vê que no caso de fulano e ciclano, eles tem a mesma média, mesma mediana e mesma moda, 8 e 8. Com as medidas de tendência central, esses caras são iguais, as notas são iguais, são o mesmo aluno, não dá pra distinguir. Pegando nosso amigo desvio padrão, DF.std, como a gente viu ali, aí sim conseguimos perceber

essa diferença, você vê que aqui essa estatística de dispersão, o desvio padrão, mostra que ciclano tem desvio padrão menor, ou seja, é mais constante, e a gente pode ver visualmente, no caso desses dados, a gente pode ver visualmente, num conjunto grande a gente não consegue.

[4:58] Você vê que a variação dele está entre 7, 8, naquela casa bem próximo da média, não varia longe da média, diferente do fulano, e nosso desvio padrão de fulano mostra isso, 2.3, varia 2.3 da média, conseguimos perceber isso olhando os dados no olhometro porque é muito pequeno, legal?

[5:21] Pessoal, medidas de dispersão, era isso que eu queria mostrar.

[5:24] Próximo vídeo, eu vou mostrar pra vocês, vou entregar pra vocês o notebook em branco já com o roteirinho pra você seguir, pra fazer as análises descritivas utilizando tudo que a gente aprendeu no nosso curso, e depois a gente faz uma correçãozinha legal, mas o foco aí nos exercícios e no próximo vídeo mostro pra vocês o notebook, beleza?

[5:45] Até o próximo vídeo, abraço.