

Conhecendo a distribuição normal

Transcrição

[0:00] Legal, galera, continuando com distribuições de probabilidade, vou lançar mais um probleminha. Em um estudo sobre alturas dos moradores de uma cidade, verificou-se que o conjunto de dados segue uma distribuição aproximadamente normal.

[0:15] Nos deu uma informação forte. Com média de 1 metro e 70, e desvio padrão 0.1.

[0:22] Legal, com essas informações, obtenha o seguinte conjunto de probabilidades, aí peço para vocês três probabilidades diferentes, a primeira, probabilidade de uma pessoa selecionada ao acaso ter menos de um metro e 80, ou ter entre um metro e 60 e um metro e 80, e o terceiro, ter mais de um metro e 90.

[0:42] O próprio problema disse qual tipo de distribuição temos que aplicar aqui, essa variável, altura, temos no nosso dataset também, ela também tem uma distribuição aproximadamente normal porque eu criei ela assim, depois vamos, no nosso projeto, mexer com ela e trabalhar exatamente com ela.

[0:59] Então, vamos estudar a distribuição normal, para gente conseguir resolver esse tipo de problema aqui, é a distribuição mais importante estatística, testes paramétricos assumem que a variável que estamos estudando segue aproximadamente uma distribuição normal.

[1:20] Aqui tem o teorema do limite central, que é um teorema muito importante em estatística, ele nos diz que com o aumento da amostra, a distribuição das médias amostrais tende a se aproximar de uma distribuição normal também, que é justamente essa curvinha em formato de sino aqui, é uma distribuição contínua, deixei um texto explicando tudo para vocês, com gráficos, fórmulas, características básicas da distribuição, então, se você precisar fazer uma consulta, revendo, estudar, pode fazer pelo seu notebook.

[1:55] Ela é uma distribuição de frequência de variável quantitativa contínua, apresenta essa forma de sino, simétrica em relação a sua média, ok, a gente falou sobre simetria, vimos justamente esse gráfico aqui, simétrica quer dizer que partindo da média, esse lado é igual a esse, tem o mesmo tamanho.

[2:21] aqui, algumas características importantes da distribuição normal, vou citar algumas, é simétrica em torno da média.

[2:30] A área sobre a curva corresponde a proporção de 100%, por causa da probabilidade, vamos falar sobre isso.

[2:38] As medidas da tendência central, vimos no curso anterior, a distribuição é simétrica, todas elas são iguais, é o que está dizendo aqui, média, mediana e moda são iguais nesse caso.

[2:50] Os extremos da curva tendem ao infinito, positivo e negativo, ambas, teoricamente, não tocam o eixo X, essa curvinha, está dizendo que tende ao infinito para cá e para lá, e teoricamente, essa curvinha azul não toca esse eixo aqui, jamais.

[3:12] O desvio padrão define o achatamento da curva, quanto maior o desvio padrão, mais achatada, mais grossa, mais larga é essa curva.

[3:22] A distribuição é definida exclusivamente pela sua média e pelo seu desvio padrão, aqui a fórmula da densidade de probabilidade aqui embaixo, temos constantes, e a gente tem aqui o desvio padrão, sigma, e o μ aqui, a média e o sigma novamente, definido por essas variáveis.

[3:39] A probabilidade sempre será igual à área sob a curva, vamos ver isso agora.

[3:48] Mostrando aqui a função, temos X , a variável normalmente distribuída, o σ é o desvio padrão e o μ é a média, a coisa da área que acabei de falar, probabilidade, obtida a partir da área embaixo desta curva delimitada pelos limites inferior e superior.

[4:05] O problema é que essa área é calculada da seguinte forma, essa fórmula medonha usando integrais definidas.

[4:14] Não se preocupem, a gente não vai ter que calcular integrais aqui, está fora do escopo do nosso curso, eu vou mostrar para vocês uma forma de calcular isso mais fácil utilizando Python, da mesma forma que fiz com as outras, e vamos fazer um cálculo manual usando tabelas padronizadas que vou mostrar no próximo vídeo como funciona, isso é basicamente para você entender uma coisa teórica, aqui é a integral, com essa integral definida, consigo calcular essa área sob essa curva dentro desses dois limites, essa área é o valor da probabilidade, o que eu quero lá em cima, por exemplo, é aqui, no caso, aqui, mais simples, a probabilidade do cara estar entre um metro e 60 e um metro e 80, esse limite é o inferior, 1.60, e 1.80 o superior.

[5:06] Vamos supor que seja aqui e aqui, quero saber justamente o valor da área, o valor da área é a probabilidade, ok? É isso que vamos calcular aqui, tem a fórmula, a mesma da densidade em cima, só que integral, representando o cálculo da probabilidade dentro desse limite, no próximo vídeo vou falar das tabelas padronizadas, uma forma manual de obtermos esses valores sem ter que calcular integrais definidas, legal? Vejo vocês no próximo vídeo, abraço.