

Distribuição t de Student

Transcrição

[0:00] Legal. Agora a gente já aprendeu a fazer um teste Z bicaudal. A partir de agora a gente vai começar a trabalhar para fazer um teste T agora, só que unicaudal.

[0:09] Unicaudal superior e unicaudal inferior. Eu vou fazer um deles e, para exercício, a gente vai treinar os outros. A ideia é a mesma, só muda o local na cauda da distribuição. Ok?

[0:20] Vamos ler o probleminha primeiro, como a gente sempre lê. "Um famoso fabricante de refrigerantes alega que uma lata de 350 ml de seu principal produto contém, no máximo, 37 gramas de açúcar."

[0:32] Eu vi "no máximo", então não passa. "Essa alegação nos leva a entender que a quantidade média de açúcar numa lata de refrigerante deve ser igual ou menor 37 gramas". Foi o que ele falou, no máximo 37, perfeito?

[0:48] "Um consumidor desconfiado e com conhecimentos em inferência estatística", que é a gente, "resolve testar a alegação do fabricante e seleciona, aleatoriamente, em um conjunto de estabelecimentos distintos."

[0:59] Ele passeou pela cidade dele, foi em vários estabelecimentos. Comprou 25 latinhas, fez uma amostra de 25, está aqui, desse refrigerante em questão.

[1:10] "Utilizando o equipamento correto o consumidor obteve as quantidades de açúcar em todas as 25 latas de sua amostra". Perfeito.

[1:17] "Assumindo que essa população se distribua aproximadamente como uma normal", dado do problema, "e considerando um nível de significância de 5%, é possível aceitar como válida a alegação do fabricante?"

[1:30] A gente vai testar isso daqui agora. Seria um teste unicaudal. [1:35] A primeira coisa que eu quero mostrar, antes de a gente falar sobre hipóteses, aqueles passos que a gente já fez nos problemas nos vídeos anteriores, é que.

[1:44] Observe que agora a gente tem um N de 25, isso é um dado do problema. Ele tem uma amostra de 25 latinhas. A gente tem aquele passo dois, que a gente seleciona qual distribuição a gente vai utilizar.

[1:56] Ou a normal, ou a T de Student. Então vamos lá, vamos fazer o teste da nossa figurinha.

[2:02] Primeira pergunta: N é maior ou igual a 30? Não, 25. Está aqui, não, então a gente sobe aqui.

[2:08] Segunda pergunta: é possível afirmar que a população se distribui como uma normal? Vamos lá em cima no problema.

[2:15] Ele diz isso aqui: "assumindo que essa população se distribua aproximadamente como uma normal..."

[2:22] Então, sim. Respondi aqui também, podemos afirmar, sim.

[2:27] Vindo aqui para baixo, a terceira pergunta: Sigma é conhecido? Ele falou alguma coisa sobre isso no problema, sobre o desvio padrão da população?

[2:35] Não, não falou nada sobre isso. O que a gente vai ter que fazer? Pegar essa amostra de 25 latinhas, calcular o desvio padrão dela e utilizar no nosso problema.

[2:44] Está aqui: o desvio padrão populacional é conhecido? Não. Então, essa pergunta aqui, não, a gente cai exatamente nessa formulazinha aqui que, lembra, é a fórmula do intervalo de confiança.

[2:57] Mas não se preocupe, a ideia é a gente ver que distribuição a gente vai utilizar e se a gente vai utilizar Sigma ou se vai utilizar S.

[3:05] Lógico, a gente não tem o Sigma, vamos usar o S, e a gente vai utilizar a distribuição T de Student. Então vamos lá.

[3:13] A gente ainda não falou da T de Student. Eu deixei aqui, "Construindo a tabela t de Student", como eu fiz lá com a normal quando a gente estudou a distribuição normal.

[3:21] Todo final de livro de estatística tem uma tabelinha com a distribuição T de Student, a distribuição normal, entre outras. A gente, nesse nosso treinamento, a gente vai conhecer mais um.

[3:34] Então eu tenho aqui um macetinho para você construir. Se você quiser tirar daqui e botar em outro Notebook é só copiar tudo. Eu já importei o Pandas, já importei tudo que é necessário para construir essa tabela.

[3:43] Então copia e cola onde você quiser, executa, ele vai rodar e vai gerar essa tabelinha aqui. A gente já vai falar sobre esse cara aqui, que é o `t_student.ppf`.

[3:53] Que é, basicamente, a mesma coisa que a gente faz com a normal. PPF, ESF, aquelas funcionalidades que eu passei para vocês.

[4:00] Tem aqui a ajuda na documentação, está ali o link para ela, com todos os métodos que a gente vai utilizar, o PPF, PDF, e por aí vai, então a gente não precisa se preocupar com isso agora.

[4:14] A preocupação é a gente entender aqui, mais ou menos, como a gente obtém as informações dessa tabela.

[4:21] Essa distribuição é bem próxima da normal, ela parece muito com a normal. Aquela curva em sino, ela também é simétrica, ela só tem as caudas um pouco mais largas.

[4:31] E ela depende, basicamente, dos graus de liberdade da distribuição que a gente está estudando. Que é o que? N a gente já sabe o que é, é o número de registros dentro da nossa amostra.

[4:42] É o número de observações, menos um. Está aqui, grau de liberdade, N menos um, são essas coluninhas aqui.

[4:49] Eu fiz até o 30. A gente está estudando até o 30, mas você pode fazer mais. Quanto maior o grau de liberdade, mais essa distribuição se aproxima de uma normal.

[4:58] E a gente não precisa usar T de Student, a gente usa a normal. Perfeito? Então até 30 a gente usa a T. Por isso que eu criei T.

[5:07] Mas como é que eu vou visualizar isso daqui? Aqui embaixo tem uma figura, vou descobrir a figura aqui, onde essa área hachurada aqui é o Alfa, a probabilidade.

[5:18] A área aqui representa probabilidade. Eu quero descobrir, dada uma probabilidade, o Alfa que a gente tem lá, a significância, vamos dizer assim, e eu quero saber qual é esse t_{α} aqui.

[5:29] Que divide a minha distribuição nessas duas partes. O Alfa e o um menos Alfa. Se ela for bicaudal, aqui vai ser o T e aqui vai ser o menos T, porque ela é simétrica com a normal, tem o zero aqui no meio.

[5:45] O que eu faço aqui? Um exemplo prático. Aqui em cima tem os Alfas para um teste unicaudal, esse debaixo aqui, e para um teste bicaudal. Então, um exemplo prático aqui.

[5:59] Por exemplo, se eu tenho uma amostra de tamanho seis, eu tenho um nível de significância de 5% e o meu teste é unicaudal, qual é o T que eu preciso selecionar, esse têzinho aqui, para eu ter os 5%?

[6:17] Como ele é unicaudal, ele vai ser nesse formato aqui. Ter 5% para cá e 95% para lá, dado o grau de liberdade, o N, seis.

[6:25] Como eu disse, os graus de liberdade são equivalentes a N menos um, então como eu tenho o N seis, seis menos um, cinco, então é esse carinha aqui.

[6:35] Eu disse que o teste era unicaudal com significância de 5%. Está aqui, Alfa 0,05, 5%. Perfeito? Então unicaudal, esse cara aqui, cinco, o meu têzinho é 2,01505. Perfeito?

[6:50] E se o teste fosse bicaudal? Vamos aqui em cima. Mesma coisa. N, seis; seis menos um, cinco. Essa linha aqui. Se ele fosse bicaudal, teria que olhar na coluna de cima.

[7:00] 0,05, que é o Alfa, eu venho aqui e vejo que o T agora é 2,57058. Perfeito? É assim que a gente faz.

[7:12] Logicamente, como na normal a gente tem funcionalidades aqui no Python que a gente já te dá o grau de liberdade, dá o Alfa e ele joga para a gente o T.

[7:20] E o contrário também acontece. A gente vai ver isso mais para frente. Nesse vídeo é isso que eu queria mostrar. Eu queria mostrar para vocês a distribuição T de Student.

[7:29] E no próximo, a gente resolve esse problema que a gente já leu aqui, beleza? Vejo você lá.