**02 - Distribuição t de Student**

[0:00] Já aprendemos a fazer um teste Z bicaudal. Agora vamos começar a trabalhar para fazer um teste T, só que unicaudal.

[0:09] Unicaudal superior e unicaudal inferior. Eu vou fazer um deles e, no exercício, vamos treinar os outros. A ideia é a mesma, só muda o local na cauda da distribuição. Ok?

[0:20] Vamos ler o problema primeiro, como sempre fazemos. "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 li "no máximo", então, não passa. "Esta alegação nos leva a entender que a quantidade média de açúcar em uma lata de refrigerante deve ser igual ou menor que 37 gramas". Foi o que ele falou, no máximo 37.

[0:48] "Um consumidor desconfiado e com conhecimentos em inferência estatística 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, desse refrigerante em questão.

amostra = [37.27, 36.42, 34.84, 34.60, 37.49,

36.53, 35.49, 36.90, 34.52, 37.30,

34.99, 36.55, 36.29, 36.06, 37.42,

34.47, 36.70, 35.86, 36.80, 36.92,

37.04, 36.39, 37.32, 36.64, 35.45]

amostra = pd.DataFrame(amostra, columns=['Amostra'])

amostraCOPIAR CÓDIGO

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

[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 agora. Seria um teste unicaudal. A primeira coisa que eu quero mostrar, antes de falarmos sobre hipóteses, são aqueles passos que a gente já fez nos vídeos anteriores.

| **n é igual ou maior a 30?** |  |  |  |
| --- | --- | --- | --- |
| Para reposta **"sim"** | O desvio padrão (sigma) é conhecido? | **Se sim**, então a média é mais ou menos igual a z vezes s sobre raiz de n | **Se não**, então a média é mais ou menos igual a z vezes o desvio padrão sobre raiz de n |
| Para resposta **"não"** | É possível afirmar que a população se distribui como uma normal? | **Se sim**, o desvio padrão é conhecido? **Sim:** a média é mais ou menos z vezes o desvio padrão sobre raiz de n; **Não:** média é mais ou menos t vezes s sobre raiz de n | **Se não**, aumente o tamanho da amostra ou utilize testes não-paramétricos |

[1:44] Observe que agora temos um n de 25, isso é um dado do problema. Ele tem uma amostra de 25 latinhas. Temos 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.

[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] 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, ao responder essa pergunta, vamos para a fórmula do intervalo de confiança.

[2:57] Mas não se preocupe, a ideia é verificar qual distribuição vamos utilizar, Sigma ou S.

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

[3:13] Ainda não falamos 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. No nosso treinamento, vamos conhecer mais um.

[3:34] Então, eu vou ensinar algo para facilitar construir. Se você quiser tirar daqui e botar em outro Notebook é só copiar tudo. Eu já importei o Pandas e tudo que é necessário para construir a tabela.

import pandas as pd

from scipy.stats import t as t\_student

tabela\_t\_student = pd.DataFrame(

[],

index = [i for i in range(1,31)],

columns = [i / 100 for i in range(10, 0, -1)]

)

for index in tabela\_t\_student.index:

for colum in tabela\_t\_student.columns:

tabela\_t\_student.loc[index, colum] = t\_student.ppf(1 -float(colum) / 2, index)

index = [('Graus de Liberdade (n - 1)', i) for i in range(1, 31)]

tabela\_t\_student.index = pd.MultiIndex.from\_tuples(index)

columns = [("{0:0.3f}".format(i/ 100), "{0:0.3f}".format(i/ 100 / 2)) for i in range(10, 0, -1)]

tabela\_t\_student.columns = pd.MultiIndex.from\_tuples(columns)

tabela\_t\_student.rename\_axis(['Bicaudal', 'Unicaudal'], axis=1, inplace = True)

tabela\_t\_student

COPIAR CÓDIGO

[3:43] Então copia e cola onde você quiser, executa, ele vai rodar e vai gerar essa tabelinha.

|  | **Bicaudal** | **0.100** | **0.090** | **0.80** | **0.70** | **0.60** | **0.50** |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Unicaudal | 0.050 | 0.045 | 0.040 | 0.035 | 0.030 | 0.025 |
| Graus de Liberdade (n-1) | 1 | 6.31375 | 7.02637 | 7.91582 | 9.05789 | 10.5789 | 12.7062 |
|  | 2 | 2.91999 | 3.10398 | 3.31976 | 3.57825 | 3.89643 | 4.30265 |
|  | 3 | 2.35336 | 2.47081 | 2.60543 | 2.7626 | 2.95051 | 3.18245 |
|  | 4 | 2.13185 | 2.2261 | 2.33287 | 2.45589 | 2.60076 | 2.77645 |
|  | 5 | 2.01505 | 2.09784 | 2.19096 | 2.29739 | 2.42158 | 2.57058 |

[3:54] Já comentarei sobre o t\_student.ppf, que é, basicamente, a mesma coisa que a gente faz com a normal. PPF, ESF, aquelas funcionalidades que eu passei para vocês.

[4:00] Temos a ajuda na documentação, [disponível neste link](https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.t.html), com todos os métodos que vamos utilizar, o PPF, PDF, e por aí vai, então a gente não precisa se preocupar com isso agora.

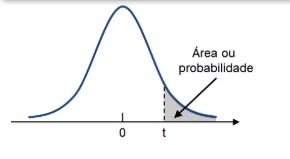
[4:14] A preocupação é entender, mais ou menos, como obter as informações dessa tabela.

[4:21] Essa distribuição é bem próxima da normal, ela parece muito com a normal. Aquela curva em sino, ela também é simétrica, 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, que é o número de registros dentro da nossa amostra, é o número de observações, menos um. Está aqui: Graus de liberdade (n-1) e, em seguida, as colunas.

[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. E a gente não precisa usar T de Student, a gente usa a normal. Então até 30 a gente usa a T. Por isso que eu criei T.

[5:07] Mas como é que eu vou visualizar isso? Aqui embaixo tem uma figura, com uma área hachurada em cinza que é o Alfa, a probabilidade.



[5:18] Eu quero descobrir, dada uma probabilidade - o Alfa, a significância, vamos dizer assim - qual é o valor de t, que divide a distribuição em duas partes: o Alfa e o 1 menos Alfa. Se ela for bicaudal, na cauda do lado direito teremos o t e do lado esquerdo -t, porque ela é simétrica como a normal e tem o zero no meio.

[5:45] Um exemplo prático. Na parte de cima da tabela temos os alfas para um teste unicaudal e para um teste bicaudal. Então, um exemplo prático.

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

[6:17] Como ele é unicaudal, ele vai ser representado no formato da imagem apresentada acima. Terá 5% para o lado direito, a partir da linha pontilhada, e 95% para o lado esquerdo, dado o grau de liberdade, o n = 6.

[6:25] Como eu disse, os graus de liberdade são equivalentes a n-1, então, como eu tenho o n = 6, 6-1 = 5. Então, o t é 2.01505.

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

[6:50] E se o teste fosse bicaudal? Vamos aqui em cima. Mesma coisa: n = 6; 6 - 1 = 5. Se ele fosse bicaudal, teria que olhar na coluna de cima. Então, 0,05, que é o Alfa, e o t agora é 2,57058. É assim que a gente faz.

[7:12] Logicamente, como na normal,temos funcionalidades no Python que, ao escrevermos o grau de liberdade e o Alfa nos apresentam o t. E o contrário também acontece, vamos discutir mais sobre isso nos próximos vídeos.

[7:25] Nesse vídeo é isso que eu queria mostrar: a distribuição t de Student.

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