**04 - Teste unicaudal**

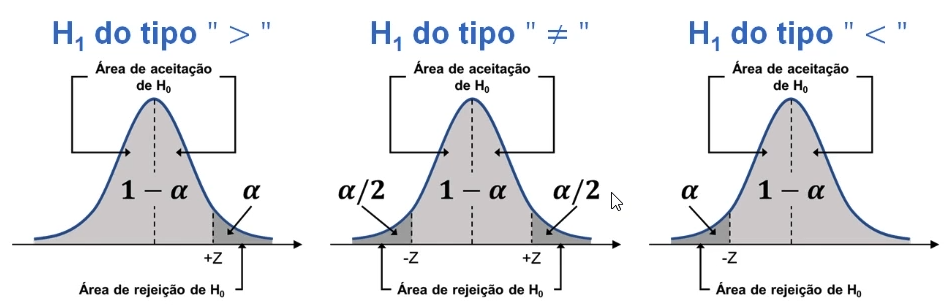
[0:00] Continuando com o teste unicaudal, já conhecemos a distribuição t de Student, que é o que a gente vai utilizar para resolver o problema que lemos no vídeo anterior. É aquele problema de que o fabricante alega que determinado produto dele tem, no máximo, 37 gramas de açúcar. Não passa disso.

[0:20] Ou seja, ele já deu a hipótese nula, a igualdade está ali. E reparem, também está no material de estudo de vocês que, geralmente, a hipótese que estamos testando fica configurada como a hipótese alternativa. No caso, eu estou testando se a alegação do fabricante é verdade ou mentira. Eu acho que tem mais de 37 gramas. É isso que eu vou testar.

[0:44] A tabela está aqui e vamos continuar com o teste unicaudal.

|  | **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 |

Aqui os tipos de testes. No caso, H1, que é a hipótese alternativa, que é do tipo maior que, a gente tem um teste unicaudal superior. Onde a área de rejeição está aqui na cauda do lado direito. Rejeição de H0. E a área de aceitação é um menos Alfa.



[1:07] O outro tipo é o H1 do tipo menor que. Onde a área de rejeição está aqui na cauda do lado esquerdo. Aqui o Alfa, depois o 1 - Alfa, que é a área de aceitação de H0.

[1:22] Esse tipo de teste verifica as variáveis em relação a um piso e a um teto. É importante analisar justamente esses tópicos que aparecem no nosso problema.

[1:31] O fabricante está alegando que é no máximo 37, então o problema está em passar daquele valor. Aqui no outro é o contrário. Perfeito? Então vamos lá.

[1:41] Os dados do problema estão aqui, do mesmo jeito que a gente fez no outro. Então, vamos rodar. Temos as 25 latinhas selecionadas no supermercado para as medições de açúcar.

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]

COPIAR CÓDIGO

[1:55] Semelhante ao outro problema, eu coloquei essa amostra dentro de um Dataframe. Estão aqui as informações, perfeito.

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

amostra

Amostra

0 37.27

1 36.42

2 34.84

3 34.60

4 37.49

5 36.53

6 35.49

7 36.90

8 34.52

9 37.30

10 34.99

11 36.55

12 36.29

13 36.06

14 37.42

15 34.47

16 36.70

17 35.86

18 36.80

19 36.92

20 37.04

21 36.39

22 37.32

23 36.64

24 35.45

COPIAR CÓDIGO

Eu vou calcular a média amostral, que é 36,25.

media\_amostra = amostra.mean()[0]

media\_amostra

36.250400000000006COPIAR CÓDIGO

[2:06] Também vou calcular o desvio padrão da amostra, porque não foi dado o desvio padrão da população. Tenho que pegar a amostra e calcular o desvio padrão para obter a estatística de teste.

desvio\_padrão\_amostra = amostra.std()[0]

desvio\_padrao\_amostra

0.9667535018469453COPIAR CÓDIGO

[2:15] O procedimento é basicamente o mesmo que a gente fez antes. Novamente, a média aqui é 37, é o que o fabricante está alegando e o que a gente está testando.

[2:28] Significância é 5%. Confiança é 1 menos a significância. O n é 25 latinhas que foram selecionadas. Agora vamos utilizar os graus de liberdade, porque estamos utilizando o t de Student.

[2:40] A gente sabe que grau de liberdade é igual a n - 1. Ou seja, 25 menos um, 24.

media = 37

significancia = 0.05

confianca = 1 - significancia

n = 25

graus\_de\_liberdades = n - 1COPIAR CÓDIGO

[2:48] Rodamos todo esse trecho de código. Maravilha, tudo pronto. As hipóteses a gente praticamente já falou. O fabricante está falando qu refrigerante dele tem no máximo 37. Ou seja, pode ter um pouco menos, mas não passa de 37, segundo ele. Ele está afirmando isso. O H1 que a gente está testando aqui é que a média é maior que 37.

[3:08] Vou testar essa minha amostra a um nível de significância de 5% e vou verificar se eu consigo resolver esse meu problema, passar por essas hipóteses.

| **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 |

[3:16] Passo dois, a gente fez no programa anterior e chegamos à conclusão de que o n é maior que 30. Temos a informação no problema de que se distribui como uma normal.

[3:27] A gente não conhece o Sigma, então chegamos a esse ponto onde temos que usar o t, que é a estatística que vem do t de Student, e o s, que é o desvio padrão da amostra.

[3:37] Está aqui tudo certinho. Vamos lá. Vamos achar o valor crítico ali, o t alfa.

[3:44] Alfa por quê? Antes a gente fez o z alfa sobre dois porque era bicaudal, a gente tem que dividir o alfa em duas partes, é simétrica.

[3:51] Agora não, o alfa está de um lado só, por isso eu vou chamar de t, porque a gente está usando o t de Student, alfa.

[3:57] Isso aqui já foi importado num vídeo anterior, mas eu vou deixar aqui para ficar bem claro: from scipy.stats import t as t\_student.

[4:08] Essa que é a função, a t. Eu não vou usá-la, porque eu vou chamar a minha estatística de t e se eu fizer isso, vai ter conflito entre a funcionalidade do Python e a minha variável.

[4:18] Então eu dei um apelido para ela, as t\_student. Toda vez que eu chamar essa função eu vou ter que chamar por esse nome. Perfeito? É só isso.

[4:26] Já deixei pronta parte da tabela que a gente construiu no vídeo anterior, onde temos justamente os graus de liberdade que eu estou interessado.

|  | **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) | 23 | 1.71387 | 1.76667 | 1.83157 | 1.90031 | 1.97825 | 2.06866 |
|  | 24 | 1.71088 | 1.76667 | 1.82805 | 1.8646 | 1.97399 | 2.0639 |
|  | 25 | 1.70814 | 1.76371 | 1.82483 | 1.89293 | 1.9701 | 2.05954 |

[4:37] Ou seja, 24 é o grau de liberdade que eu estou interessado. Eu tenho um teste unicaudal. E o que eu tenho que fazer? Verificar, na minha tabela, a linha unicaudal. Qual é o nível de significância? 5%, isto é, 0,05.

[5:02] O meu grau de liberdade está aqui, então o meu t vai ser 1,71088. Perfeito? É isso. Na tabela eu já resolvi esse problema. Está aqui: 1,71088.

[5:13] Mas tem como a gente fazer isso também utilizando o nosso amigo Python, com t\_alpha, lembrando que não tem mais o alfa sobre dois. Vai ser igual. Aí eu chamo t\_student, do mesmo jeito que a gente faz com a normal, .ppf. Só que aqui eu passo a confiança. Já vou explicar porque. E também graus\_de\_liberdade que a gente já rodou.

t\_alpha = t\_student.ppf(confianca, graus\_de\_liberdade)

COPIAR CÓDIGO

[5:43] Por que a confiança? Verifiquem aqui. Essa função vai escolher o t, mas ela vai me passar o quê?

[5:52] Eu tenho que passar essa probabilidade, como a normal. Ela vem desse ponto, da linha pontilhada, até o menos infinito, por isso que eu estou passando a 95%. Aqui, já está pronto, confiança 95%.

[6:04] Desse ponto até aqui, eu tenho que passar a probabilidade com o grau de liberdade que eu estou assumindo, para que ele me retorne 1,71.

[6:14] Então, vamos lá. t alfa é igual a 1,71. E já está desenhado aqui, já definimos a área de rejeição e a área de aceitação, ok?

t\_alpha = t\_student.ppf(confianca, grau\_de\_liberdade)

t\_alpha

1.7108820799094275COPIAR CÓDIGO

[6:26] Cálculos da estatística de teste: exatamente igual fizemos no vídeo anterior. Só que aqui eu vou chamar ele de t. E t vai ser igual a, primeiro, media\_amostra menos a media.

t = (media\_amostra = media))

/ ()COPIAR CÓDIGO

[6:46] Depois, desvio\_amostra, dividido por np.sqrt() - função para extrair a raiz quadrada de um número, recebe n.

t = (media\_amostra = media))

/ (desvio\_padrao\_amostra / np.sqrt(n))

-3.876893119952045COPIAR CÓDIGO

[7:03] Eu vou ter um t de menos 3,876. Eu arredondei isso para menos 3,88. Já percebemos que ele está na área de aceitação de H0.

[7:20] Ou seja, o nosso problema de teste de hipóteses já está resolvido visualmente. Mas vamos supor que a gente não tenha esse recurso gráfico e queremos ter certeza.

[7:28] Vamos até o unicaudal superior, esse é o passo cinco. Unicaudal superior; aqui onde estão as hipóteses, igual a nossa. A gente obteve uma determinada estatística.

[7:41] Tirando o Sigma, usando o s. E a gente tem que vir aqui, esse é o crítico aqui para o t: rejeitar o H0 se t, que é a estatística que a gente calculou, esse - 3,88, é maior ou igual. A pergunta é: t, você é maior ou igual a t Alfa?

t >= t)\_apha

False

COPIAR CÓDIGO

[8:09] Então vamos rodar isso daqui. Ele vai dizer que não, que é falso. Ou seja, eu não posso rejeitar H0. Já concluo que o fabricante não está mentindo.

[8:19] Deixei aqui embaixo escrito que: "com um nível de confiança de 95% não podemos rejeitar H0, ou seja, a alegação do fabricante é verdadeira."

[8:27] Lógico, baseado na amostra que eu selecionei e nesse nível de confiança, ou nível de significância. No próximo vídeo, eu vou mostrar para você, como a gente fez antes, o P valor.

[8:41] E também uma forma de calcular esse Ttest. A gente está fazendo agora o Ttest, a gente fez o Ztest antes. Calcular o Ttest utilizando ferramentas do Statsmodels.

[8:51] Até lá.