**04 - Entendendo p-valor**

[0:00] Pessoal, continuando o exercício que a gente estava fazendo, vamos falar agora do P valor. Se você já estudou estatística, provavelmente você já ouviu falar dele, e eu vou mostrar para vocês como calcular e como interpretar.

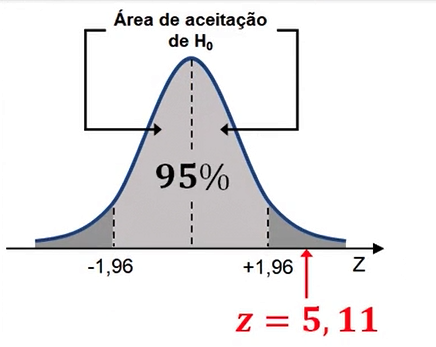
[0:10] É uma coisa bem simples. Voltando ao nosso quadro do vídeo anterior, onde temos as regras de rejeição de H0, uma delas é o P valor.

[0:20] O "t" a gente vai estudar no próximo vídeo. Nós rejeitamos H0 se P valor for menor do que Alfa. Alfa é o que? É o nosso nível de significância. Geralmente, a gente ouve falar: é menor que 0,5.

[0:33] Isso quando o nível de significância é de 5%. Se for 1%, ele tem que ser menor do que 0,01. Perfeito? Então preste atenção só nesse detalhe.

[0:43] O que é o P valor? Lembra que a gente calculou a estatística de teste? Ela chegou e caiu em 5,11, ok? O P valor é justamente a probabilidade, a área calculada do ponto z = 5,11. O risco pontilhado (+1,96) é o nível de significância, alfa sobre dois.

[1:01] A partir deste pontilhado, nós temos uma área (a cauda direita), uma probabilidade. O P valor é justamente a probabilidade do valor crítico, à frente, à direita, nesta área. O que está querendo dizer aquele teste? Que a área do ponto crítico for menor que a de significância, quer dizer que minha estatística está dentro da região de rejeição. Ou seja, quando o P valor é menor do que Alfa, eu rejeito H0. Perfeito?



[1:35] É justamente isso, apenas outra forma, mais simples, porque todo software estatístico vai soltar para você nos testes o P valor, e aí você já tem na sua cabeça que essa regra é a de rejeição da hipótese nula.

[1:49] Então, vamos lá. Vamos calcular o P valor e vamos ver se ele concorda também com essa decisão que a gente tomou no vídeo anterior.

[2:01] Uma das primeiras formas é o que? Eu vou chamar de p\_valor mesmo. No caso do teste bicaudal, eu tenho que multiplicar por dois, pois temos duas caudas.

[2:24] Duas vezes. Eu abro e fecho um parênteses. A primeira forma é a seguinte: 1 - norm.cdf(z). Aí eu vou ter um p\_valor.

p\_valor = 2 \* (1 - norm.cdf(z))

p\_valor

3.2291031715203644e-07COPIAR CÓDIGO

[2:42] O meu P valor é bem pequeno, não é? Tem um monte de zeros até chegar ao 3,22. Aqui eu já deixei uma situação formada, que é aquela pergunta: o P valor é menor ou igual a significância? Que é 0,05?

p\_valor <= siginificancia

TrueCOPIAR CÓDIGO

[2:57] Eu já consigo responder que é. Se é, eu rejeito H0, foi a conclusão que chegamos anteriormente. Então é isso.

[3:07] A outra forma que eu queria mostrar para você, utilizando também o mesmo recurso do Norm, é duas vezes também, só no teste bicaudal, preste atenção nisso.

[3:20] Duas vezes. E aqui eu vou chamar de Norm também, ponto. Aqui é o sf e eu passo para ele o z. O z, lembrando, que é justamente o cálculo da probabilidade do ponto crítico, 5,11.

p\_valor = 2 \* (nome.sf(z))

p\_valor

3.2291031715203644e-07COPIAR CÓDIGO

[3:44] Lógico, tem uma diferencinha aqui. A documentação diz que ele, usando o SF, é um pouco mais preciso.

[3:52] Então, você pode escolher porque não vai atrapalhar a sua decisão, a precisão.

[4:03] Então, vamos lá. Fizemos o teste, já estudamos o P valor, já entendemos como calcular. Como é que a gente faz isso utilizando o Python, de maneira mais fácil, para não ter que realizar todos aqueles passos? Já que a gente já aprendeu a fazer aquilo.

[4:14] E eu sou um pesquisador experiente, não estou precisando fazer uma prova escrita, como é que eu faço isso usando o Python? Vamos lá.

[4:20] Primeira coisa, eu vou fazer usando o método statsmodels.stats.weightstats.ztest. Eu vou usar o ztest, que é justamente o que estamos fazendo, um teste Z. Portanto, faremos o código a seguir.

from statsmodels.stats.weightstats import ztestCOPIAR CÓDIGO

[04:47] É isso. Em seguida, apertamos "Shift + Enter", rodou. [Este link](https://www.statsmodels.org/stable/generated/statsmodels.stats.weightstats.ztest.html) abre a janela do StatsModels.

[4:51] Eu já vou mostrar as outras. Também mostrarei uma segunda opção para realizar o ztest, usando o StatsModels.

[4:58] E o que eu vou fazer? Chamo o ztest e passo para ele o x1 igual a amostra, que é uma amostra que a gente selecionou, está lá em cima aquele Dataframe.

[5:12] E eu passo para ele value que é a média que eu estou testando, aqueles 500 ml. Perfeito? Ele vai jogar para a gente o quê? Duas saídas.

ztest(x1 = amostra, value = media)

(array([5.10955978]), array([3.2991037e-07]))COPIAR CÓDIGO

[5:23] Uma vai ser a estatística de teste, que é o 5,11 que a gente calculou, e o outro é o P valor, muito semelhante ao anterior.

[5:40] Isso é o Output do teste. Outro teste que eu quero mostrar para vocês é o DescrStatsW. Vou copiar ele aqui, para facilitar a nossa vida, e vou colar. Acessando [este link](https://www.statsmodels.org/dev/generated/statsmodels.stats.weightstats.DescrStatsW.html), você vai abrir aquela ajuda do Statsmodels.

from statsmodels.stats.weightstats import DescrStatsWCOPIAR CÓDIGO

[6:16] Para usar essa ferramenta, eu tenho que, primeiro, chamar test. Vou criar um objeto do tipo DescrStatsW() e vou passar para ele a minha amostra, a amostra aqui para dentro.

test DescrStatsW(amostra)COPIAR CÓDIGO

[6:33] Com esse objeto teste, eu vou realizar o teste ztest\_mean. Posso copiar ele aqui.

[6:46] Pego o test.ztest\_mean(), passo para ele o value, que é igual a média.

test.stast\_mean(value = media)COPIAR CÓDIGO

[6:56] Por que eu estou passando só o value? Lembra que eu criei um objeto com a amostra? Ele já tem os meus dados. Eu vou pegar e realizar um teste nesses dados. E ele vai, similar ao de cima, colocar a estatística de teste e o meu P valor.

(array([5.10955978]), array([3.2991037e-07]))COPIAR CÓDIGO

[7:14] Se eu quiser separar estes dois, eu posso chamar de z, p\_valorw igual a test.stast\_mean(value = media).

z, p\_valor = test.stast\_mean(value = media)COPIAR CÓDIGO

[7:22] Como ele está dentro de um Array, eu vou fazer um print para o z, só que eu vou passar o zero. Então, ele vai me retornar a estatística de teste.

z, p\_valor = test.stast\_mean(value = media)

print(z[0])

5.10955978COPIAR CÓDIGO

[07:30] Eu posso fazer o mesmo para o p\_valor. Lembre-se de que ele estava dentro de um Array, eu quero pegar o elemento que estava dentro.

z, p\_valor = test.stast\_mean(value = media)

print(z[0])

print(p\_valor[0])

5.10955978

3.2991037e-07COPIAR CÓDIGO

[7:44] O valor de cima é a minha estatística e o abaixo é o meu P valor.

[7:49] Pessoal, era basicamente isso que eu queria mostrar. Já temos um ferramental forte para você utilizar, para realizar teste bicaudal.

[8:00] Só mais uma coisa antes de finalizarmos. Eu não configurei nada sobre ele ser bicaudal ou unicaudal. Por quê? Esses outros tipos de testes, que encontramos na página do DescrStats, essa hipótese Alternative pode dizer *two-sided*.

[8:14] Ou seja, ele já vem com padrão bicaudal. No próximo vídeo, vamos estudar o unicaudal e vamos alterar essa opção.

[8:22] Então, no próximo vídeo, estudaremos teste unicaudal paramétrico. Até lá.