

Resumo

Transcrição

[0:00] Moçada, maravilha. A gente está finalizando aqui o nosso curso de teste de hipóteses, então vamos dar uma passada rápida de olho no que a gente aprendeu.

[0:08] Para depois a gente partir para resolver um último exercício no Notebook que eu vou deixar de presente para você. Vamos lá.

[0:13] A gente começou, justamente, já levando um susto, fazendo um teste de normalidade, que é aquele teste onde a gente verifica se determinada variável se distribui como uma normal, coisa que é bastante importante em estatística.

[0:27] Então a gente viu o normaltest. Fizemos um teste em duas variáveis do nosso Dataset, uma delas não se distribuía como uma normal, que é a renda.

[0:36] E a outra, que é uma variável construída, é um fake, de alturas, que tem um desenho bonitinho aqui, se distribuiu como uma normal. Beleza?

[0:43] A gente viu as duas coisas inicialmente só para tomar aquele susto e ver como funciona um teste estatístico. Depois a gente viu as etapas do teste. Isso aqui é muito importante.

[0:52] Mesmo que você resolva desenvolver os testes com aquele ferramental que eu mostrei para vocês do Python, que é muito mais simples. Com uma linha de código a gente já resolve o teste.

[1:02] Mas mesmo assim, mentalmente ou numa folha de papel, você tem que pelo menos definir as suas hipóteses, escolher o teste que você vai utilizar, coisas desse tipo.

[1:11] Para você poder saber também o que está acontecendo. Se você quer um teste unicaudal, um teste bicaudal, isso tudo a gente vê definindo as nossas hipóteses, que é o nosso primeiro passo aqui.

[1:20] Lembrando que o H_0 é quem contém a igualdade, coisa que a gente vem falando sempre. O outro passo, que é a escolha da distribuição que a gente vai utilizar.

[1:29] Isso a gente faz nos testes paramétricos. Deixei aqui, de presente, essa figurinha. Guarde com você, porque aqui você tem o passo a passo para escolher de forma correta a distribuição que você vai utilizar.

[1:40] No caso aqui, o Z, que é o Z test, ou T, que seria o T test.

[1:46] O próximo passo é a definição das áreas críticas no nosso gráfico aqui, área de rejeição e de aceitação da hipótese nula, isso a gente já está bem amaciado para fazer.

[1:59] A gente já começou mexendo com esse tipo de coisa lá no outro curso, onde a gente aprendeu a obter as probabilidades de distribuição normal.

[2:07] E a gente vem continuando fazendo isso até agora. E ainda vai continuar mais um pouquinho no próximo.

[2:12] Aqui, os tipos de testes que existem: o unicaudal superior; o unicaudal inferior; e o bicaudal, que é esse do meio aqui. E como é definido H_1 para cada um destes testes.

[2:24] H1, que é a hipótese alternativa. Que, de modo geral, quando a gente está realizando o teste, é justamente a hipótese que a gente está querendo testar - eu tinha dito isso para vocês.

[2:34] E a gente vai ver no exercício final um exemplo disso. O outro passo é o cálculo da estatística de teste, onde a gente vai fazer as comparações e ver se essa estatística cai na área de rejeição.

[2:46] Ou se cai na área de aceitação da hipótese nula. Também deixei lá, em cada probleminha, um quadrozinho, onde você tem os tipos de teste e também as áreas de rejeição, as estatísticas de teste.

[3:00] Eu já vou mostrar aqui. E o último passo, que é, justamente, a tomada de decisão; aceita ou rejeita a hipótese nula ou a hipótese alternativa.

[3:09] Vamos lá. Passamos para os testes paramétricos, que são aqueles testes que assumem algumas premissas sobre os parâmetros, como os parâmetros de uma população se distribui.

[3:19] E a gente conheceu o Z test, o T test. Primeiro a gente resolveu um teste bicaudal.

[3:28] Vamos voltar, vamos vir aqui no quadro. Aqueles passos que a gente fez, cálculos das estatísticas. Aqui aquele quadrozinho.

[3:35] O bicaudal é onde a hipótese nula é desse tipo aqui: a média igual a um valor ou a média é diferente de um valor? É onde não importa se ela é para cima ou para baixo e sim a igualdade.

[3:48] E os testes unicaudais, que a gente fez um deles aqui. A gente fez um unicaudal superior, a gente depois vai fazer um unicaudal inferior, no próximo vídeo.

[4:03] E está aqui tudo o que a gente precisa saber, praticamente, um resuminho de cada tipo de teste, as hipóteses que a gente tem que formular, onde as estatísticas vão estar posicionadas.

[4:15] O cálculo das estatísticas de teste e as regras de rejeição de H_0 para o valor crítico, para o valor crítico das duas variáveis, no caso o Z e o T, e para o P valor, que é sempre o mesmo para todos eles.

[4:29] Aqui, a gente fez tudo isso, decidimos o teste, tomamos uma decisão, e depois eu mostrei para você como achar um P valor, calcular o P valor, e depois sempre utilizando o ferramental do Python, que é o Statsmodels ou o Scipy.

[4:47] Aqui, seria o mesmo teste de cima, só que na versão unicaudal, a gente fez um unicaudal superior. Aprendemos também a distribuição T de Student, a consultar uma tabela da distribuição T de Student.

[5:04] Aqui, todo o teste, todos aqueles passos que a gente realizou. Aqui também aquela mesma tabelinha, isso aqui é igual ao do anterior, porque os testes são parecidos.

[5:15] E aqui também a gente fez um teste paramétrico para testar duas amostras, que é basicamente o que a gente fez no vídeo anterior, que é o Mann-Whitney, só que essa aqui é a versão paramétrica dele.

[5:27] Utilizando o teste Z ou o teste T, depende do tamanho da amostra.

[5:33] Aqui, a forma de criar as hipóteses e aqui também a estatística de teste, que tem uma diferencinha das outras, lógico.

[5:46] Está aqui a tabelinha também, própria para esse tipo de teste quando a gente está mexendo com duas amostras.

[5:53] E a gente foi para os testes não paramétricos, onde a gente conheceu o Qui-quadrado, conhecemos a tabela Qui-quadrado, como consultar uma tabela Qui-quadrado.

[6:01] Que é aquele teste que a gente testa as frequências observadas e esperadas, e toma as decisões. A gente fez uma brincadeira da moeda do juiz.

[6:10] Aqui, a forma da distribuição do Qui-quadrado. Aqui também mostrei como calcular as probabilidades, a estatística de teste, a gente fez tudo manualmente para o pessoal que tem que fazer as provas escritas.

[6:25] Aqui, aquela tabelinha que eu falei, para o Qui-quadrado. E a gente foi para o teste de Wilcoxon, que é um teste que testa duas populações, assim como aquele teste T que a gente fez, o teste Z.

[6:36] Só que aqui ele é a versão não paramétrica e esse cara aqui testa amostras dependentes.

[6:40] É quando a gente está, por exemplo no nosso exemplo aqui mesmo, verificando um grupo que está sendo submetido a um tratamento.

[6:48] A gente verifica informações deles antes do tratamento e verifica informações depois. É o mesmo grupo, por isso eles são dependentes.

[6:56] E aí todo aquele processo que a gente vem fazendo, estatística de teste, calcular. Isso aqui é um pouco mais chatinho de fazer na mão porque tem que montar aquela tabela.

[7:03] Você deve lembrar, a gente fez aquela tabela dos postos. E depois fazer as somas dos postos.

[7:09] Você vê o tamanho do nosso arquivo, ficou grande por causa desses dois problemas, o Wilcoxon e Mann-Whitney, que a gente vai ver daqui a pouco.

[7:16] Aqui, o cálculo das estatísticas. A estatística desses caras, com o N subindo ela converge para uma distribuição normal, mas não tem nada a ver com ser paramétrico, isso é estatística de teste. Perfeito?

[7:30] Não é a população, a amostra que a gente está testando, que se distribui como uma normal. Aqui os testes, aquela tabelinha também.

[7:40] O teste bicaudal que a gente realizou. Logicamente, mostrando tudo como faz no Python, de maneira simples.

[7:46] E o teste de Mann-Whitney, que foi o último que a gente fez, para duas amostras, só que independentes agora.

[7:51] 4E a gente fez para uma amostra pequenininha, que é uma das formas de aplicar um teste não paramétricos, em amostras muito pequenas, amostras menores que 30, por exemplo.

[8:02] Onde o teorema do limite central não se aplica, e a gente consegue realizar testes utilizando esses caras aqui.

[8:08] E a gente fez o mesmo teste que a gente fez lá para as duas amostras no paramétrico, fizemos aqui com o Mann-Whitney, aquela coisa da renda do homem e da mulher. Perfeito?

[8:18] Então, é isso. Aqui também um teste mais trabalhoso, porque tem que fazer todas aquelas tabelas, mas na hora de resolver com ferramentas do Python é bem simples.

[8:28] Uma linha de código aqui, mannwhitney, mulheres, homens, alternative Less. Pronto, ele já te dá o P valor, a estatística de teste e já está resolvido.

[8:37] Pessoal, é isso aqui que eu queria mostrar, só dar uma passada de olho em tudo que a gente viu.

[8:42] No próximo vídeo, eu vou apresentar para vocês o exercício, aí vocês vão resolver, logicamente. E depois a gente faz uma correção dele, eu mostro para vocês o notebook prontinho com a correção. Beleza? Até lá.