**02 - Resumo**

[0:00] Estamos finalizando o nosso curso de teste de hipóteses, então vamos revisar, rapidamente, o que aprendemos, para depois resolvermos 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, em que verificamos se determinada variável se distribui como uma normal, coisa que é bastante importante em estatística.

[0:27] Estudamos 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 se distribuiu como uma normal.

[0:43] Passamos por elas, inicialmente, só para tomar um susto e entender como funciona um teste estatístico. Depois estudamos as etapas do teste. Isso é muito importante.

[0:52] Mesmo que você resolva desenvolver os testes com o ferramental do Python, que é muito mais simples, uma linha de código já resolve o teste, mentalmente ou numa folha de papel, você tem que ao 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 ou bicaudal. Com isso tudo vamos definindo as nossas hipóteses, que é o nosso primeiro passo.

[1:20] Lembrando que o H0 é quem contem a igualdade, coisa que a gente vem falando sempre. O outro passo é a escolha da distribuição que vamos utilizar.

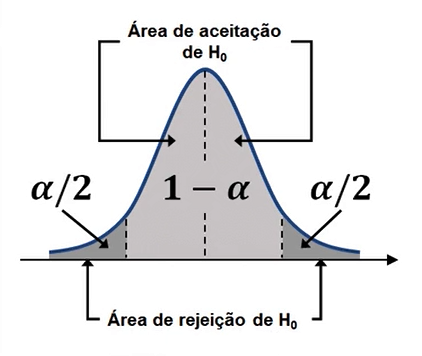
[1:29] Isso fazemos nos testes paramétricos. Deixei, de presente, essa figurinha.

| **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 |

Guarde com você, porque nela há o passo a passo para escolher de forma correta a distribuição que você vai utilizar.

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

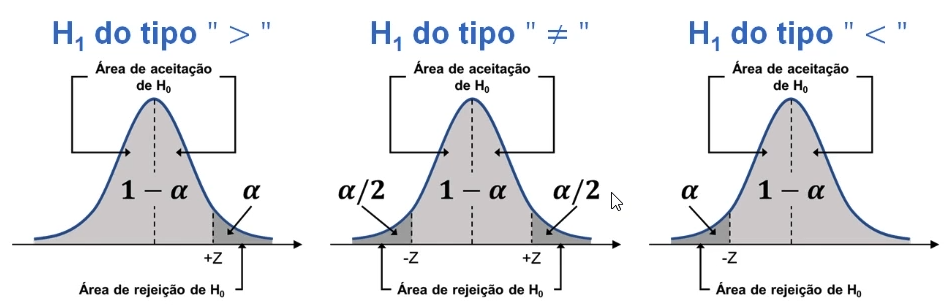
[1:46] O próximo passo é a definição das áreas críticas no nosso gráfico, área de rejeição e de aceitação da hipótese nula. Nisso já estamos bem habituados.



[1:59] Começamos mexendo com esse tipo de coisa no outro curso, onde aprendemos a obter as probabilidades de distribuição normal.

[2:07] Continuamos fazendo isso até agora e continuaremos mais um pouco no próximo.

[2:12] Passamos pelos tipos de testes que existem: o unicaudal superior; o unicaudal inferior; e o bicaudal. E como é definido H1 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 queremos testar. Eu tinha dito isso para vocês. Teremos, no exercício final, um exemplo disso.

[2:38] Outro passo é o cálculo da estatística de teste, onde faremos as comparações e descobrir se essa estatística cai na área de rejeição ou se cai na área de aceitação da hipótese nula.

[2:49] Também deixei lá, em cada problema, um "quadrozinho", onde você tem os tipos de teste e também as áreas de rejeição, as estatísticas de teste.

[3:00] E o último passo 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] Conhecemos o Z test, o t test. Primeiro, a gente resolveu um teste bicaudal.

[3:35] O bicaudal é onde a hipótese nula é desse tipo: 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 sobre os testes unicaudais, nós fizemos um deles aqui, um unicaudal superior. Depois faremos um unicaudal inferior, no próximo vídeo.

[4:03] No notebook temos tudo que precisamos saber, praticamente. Um resumo de cada tipo de teste, as hipóteses que precisamos formular e onde as estatísticas estarão posicionadas.

[4:15] O cálculo das estatísticas de teste e as regras de rejeição de H0 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] Decidimos o teste, tomamos uma decisão, e depois eu mostrei para você como achar um p valor, calcular o P valor, sempre utilizando o ferramental do Python, que é o Statsmodels ou o Scipy.

[4:47] Em seguida, estudamos algo que seria o mesmo teste de cima, só que na versão unicaudal, fizemos um unicaudal superior. Aprendemos também a distribuição t de Student, a consultar uma tabela da distribuição t de student.

[5:04] No notebook temos todos aqueles passos que a gente realizou. Aquela mesma tabelinha, igual ao do anterior, porque os testes são parecidos.

[5:15] Também fizemos um teste paramétrico para testar duas amostras. É basicamente o que a gente fez no vídeo anterior, o Mann-Whitney, só que a versão paramétrica dele, utilizando o teste z ou o teste t, depende do tamanho da amostra.

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

[5:46] Temos uma tabelinha também, própria para esse tipo de teste quando, quando estamos lidando com duas amostras.

[5:53] E passamos para os testes não paramétricos, onde conhecemos o Qui-quadrado, a tabela Qui-quadrado, como consultar uma tabela Qui-quadrado. Que é aquele teste em que testamos as frequências observadas e esperadas, e toma as decisões. Fizemos uma brincadeira da moeda do juiz.

[6:10] Estudamos a forma da distribuição do Qui-quadrado. 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] No notebook também encontraremos aquela tabelinha que eu falei, para o Qui-quadrado. E fomos para o teste de Wilcoxon, que testa duas populações, assim como aquele teste t e o teste Z, só que na versão não paramétrica e ele testa amostras dependentes.

[6:40] É quando a gente está verificando, por exemplo, um grupo que está sendo submetido a um tratamento.

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

[6:56] Tivemos todo aquele processo de estatística de teste. Isso é um pouco mais complicado de se fazer manualmente, porque precisamos montar uma tabela.

[7:03] Você deve lembrar, nós fizemos uma tabela dos postos e depois as somas dos postos.

[7:09] Nosso arquivo ficou grande por causa desses dois problemas, o Wilcoxon e Mann-Whitney, que a gente vai estudar daqui a pouco.

[7:16] Depois tivemos o cálculo das estatísticas. A estatística com o n subindo convergem para uma distribuição normal, mas não tem nada a ver com ser paramétrico, isso é estatística de teste.

[7:30] Não é a população, a amostra que estamos testando que se distribui como uma normal. Em seguida vieram os testes e uma tabela também.

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

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

[7:51] E a gente fez para uma amostra pequenininha, que é uma das formas de aplicar um teste não paramétrico, usar amostras muito pequenas, amostras menores que 30, por exemplo, onde o teorema do limite central não se aplica.

[8:08] Fizemos também o mesmo teste realizado para as duas amostras no paramétrico, mas com o Mann-Whitney. Foi aquele relacionado à renda do homem e da mulher. Perfeito?

[8:18] Então, é isso. Também um teste mais trabalhoso, porque tem que fazer muitas tabelas, mas resolver com ferramentas do Python é bem simples.

[8:28] Uma linha de código com mannwhitney, mulheres, homens, alternative e less e pronto, ele dá o p valor, a estatística de teste e já está resolvido.

[8:37] Pessoal, é isso que eu queria mostrar.

[8:42] No próximo vídeo, eu vou apresentar para vocês o exercício, vocês vão resolver, logicamente. E depois fazemos a correção. Eu mostro para vocês o notebook pronto com a correção. Até lá.