

## Teste Qui-quadrado

### Transcrição

[0:00] Legal, pessoal. A gente já começou a trabalhar com testes não paramétricos, já conhecemos a distribuição Qui-quadrado, que a gente vai utilizar para resolver esse probleminha.

[0:07] Então vamos ler o probleminha aqui e vamos resolver ele utilizando o teste, está aqui em cima, do Qui-quadrado.

[0:16] Problema: "Antes de cada partida do campeonato nacional de futebol, as moedas utilizadas pelos árbitros devem ser verificadas para se ter certeza que não são viciadas."

[0:26] "Ou seja, que não tendam para determinado resultado". Ou seja, moedas bem equilibradas, que sejam honestas.

[0:34] "Para isso um teste simples deve ser realizado antes de cada partida. Este teste consiste de lançar a moeda do jogo 50 vezes e contar as frequências de caras e coroas obtidas."

[0:45] "A tabela abaixo mostra o resultado obtido no experimento." Vamos lá. Os valores observados no experimento foram: número de caras, 17; e número de coroas, 33.

[0:54] A soma disso tudo dá 50. E o esperado, obviamente, é de que a moeda, se ela for equilibrada, ela vai ter meio a meio, 25 para um lado e 25 para o outro.

[1:02] "A um nível de significância de 5%, é possível afirmar que a moeda não é honesta, isto é, que a moeda apresenta uma probabilidade maior de cair com a face cara voltada para cima?"

[1:14] Lembra que o Qui-quadrado, a hipótese nula que é testada, é de não haver diferença entre as frequências observadas de determinado evento.

[1:25] Ou seja, aquelas que a gente testou, e as frequências realmente esperadas. O que eu estou esperando aqui é que a frequência do número de caras seja igual a frequência do número de coroas, perfeito?

[1:35] Essa vai ser a minha hipótese nula. Antes de a gente já formular as hipóteses, vamos rodar os dados do teste. Então eu criei uma lista aqui.

[1:47] Estou chamando de  $F_{\text{observada}}$ , que é a frequência observada, e nessa lista a primeira vai ser cara e a segunda coroa, então 17 e 33.

[1:56] Aqui embaixo,  $F_{\text{esperada}}$ , 25, 25, meio a meio para cada um. Significância: 0,05, 5%. Confiança: um menos a significância.

[2:09]  $K$ , que é o número de eventos possíveis. Ou seja, cara e coroa têm quantos? Dois eventos possíveis, então o nosso  $K$ , que seria o nosso  $N$  nos testes que a gente estava vendo anteriormente.

[2:21] Eu chamei de  $K$  aqui para diferenciar um pouco isso. Número de evento possíveis eu tenho o quê? Dois, cara ou coroa.

[2:29] Grau de liberdade a gente era o  $N$  menos um, agora vai ser o  $K$  menos um, ou seja, o grau de liberdade a gente tem um. Rodamos isso aqui, vamos as hipóteses.

[2:40] Os passos são basicamente os mesmos que a gente fez nos paramétricos, só que o passo dois não é necessário, então a gente não vai nem fazer o passo dois.

[2:47] A gente chamou o passo dois de outra coisa. Como eu disse, a formulação das hipóteses, eu quero testar na hipótese nula que não tenha diferença, quer dizer, que elas são iguais.

[2:56] E, aqui, a hipótese alternativa é que existe uma diferença dessas frequências, da observada para a esperada.

[3:06] Fixação do nível de significância do teste, era o passo três antes, passamos para o dois.

[3:11] Importei aquilo que a gente tinha importado no vídeo anterior para construir aquela tabela do Qui-quadrado. `from scipy.stats import chi`, ou `Qui`, chame como quiser, é Qui-quadrado.

[3:23] Aqui, eu estou chamando a tabela que a gente criou no vídeo anterior. Parte dela só, a parte que me interessa, para não ficar uma tabela muito grande, e a gente ver como consultar a tabela.

[3:34] Eu quero o quê? Um nível de significância de 5%, ou seja, um nível de confiança de 95%. Eu venho aqui no nível de confiança, que é esse cara aqui.

[3:44] Lembrando que é essa área aqui a área de aceitação de  $H_0$ . Essa tabela me dá o quê? Eu vou escolher um valor aqui e esse valor vai separar essa distribuição em duas partes.

[3:58] A parte de cá é justamente o que está mostrando aqui em cima, seria o P, perfeito? Então eu quero que tenha 95% aqui e 5% aqui. Significância, confiança.

[4:12] Com 95% de confiança eu tenho o quê? Para um grau de liberdade, duas possibilidades, dois menos um, um, eu tenho o valor de 3,8415.

[4:22] Perfeito, a gente já tem o nosso Qui-quadrado onde corta aqui. 3,8415. Como é que eu acho isso se eu não tiver uma tabela na mão?

[4:30] Vou chamar aqui de Qui-quadrado Alfa. Repara que é o mesmo procedimento que a gente vem fazendo no paramétrico, a gente vai fazer uma coisa bem parecida aqui.

[4:41] Chi quadrado, dois, não A. Alfa vai ser o quê? Chi, chamo aquela função, ponto PPF; a mesma coisa que a gente fez na normal, no T de Student.

[4:51] Aqui eu passo o quê? Confiança vírgula grau de liberdade. Só que aqui tem uma diferencinha. Lembra que a gente está trabalhando com Qui-quadrado?

[5:04] Esse aqui é o Qui, então a gente tem que elevar ele ao quadrado. É só isso.

[5:11] Lembre-se que aqui a gente está fazendo manualmente, na força, no braço, depois eu vou mostrar para você a forma de fazer tudo isso aqui simples e rápido, usando uma linha de comando do Python, perfeito?

[5:22] 3,84, como a gente tinha descoberto aqui em cima. 3,8415, esse cara aqui.

[5:29] Estatística de teste, próximo passo. Essa estatisticazinha simpática aqui, onde a gente tem um somatório até K elementos, ou seja, aquele K que a gente tem lá em cima, K igual a dois.

[5:43] Então, a gente tem duas parcelas nessa soma, é isso que está querendo dizer esse somatório aqui, vai de um a K.

[5:49] Eu botei um aqui porque em estatística a gente começa sempre a contar de um até o N. Em computação, a gente vai ter que fazer um menos um disso daqui, porque os índices começam com zero e vão até N menos um.

[6:02] É só a gente trocar. Isso aqui é só uma notação estatística.

[6:07] A formula, onde eu tenho aqui cada parcela a frequência observada, menos a frequência esperada, tudo isso elevado ao quadrado, dividido pela frequência esperada.

[6:18] Está tudo aqui bem documentadinho. Então eu vou calcular de duas formas: uma manual, para você entender essas parcelas.

[6:24] Vou chamar de Chi\_2, Qui-quadrado. Essa formula aqui em cima está querendo dizer o quê? Tem duas parcelas nessa soma, a primeira e a segunda.

[6:33] Na primeira, eu vou colocar o quê? Eu vou já abrir parênteses aqui dentro também, porque eu tenho um numerador e um denominador.

[6:42] Eu vou fazer o quê? F, eu quero o F\_observado. Como aquilo é uma lista, para eu pegar o valor dentro da lista é o quê?

[6:52] O meu primeiro índice aqui é um, mas lembra que em computação a gente começa com zero, então um quer dizer zero. Menos F\_esperada.

[7:04] Mesma coisa, o índice ali ao primeiro é zero. Isso tudo aqui elevado ao quadrado e dividido por esse cara aqui. Está escrito ali, F\_esperado, o mesmo cara.

[7:21] São duas parcelas, como eu disse, a primeira é essa. A segunda é o quê? É o outro índice. Vamos copiar esse cara aqui todo, que está aqui dentro, e vamos colocar aqui dentro.

[7:35] Só vamos trocar os índices, agora não é mais zero, é um, é o próximo valor que a gente está somando. Essa aqui é a conta que eu tenho que fazer.

[7:46] 5,12. Vou fazer de uma forma mais elegante, um pouco mais simples. Chi\_2, eu vou atribuir um zero, porque eu estou iniciando essa variável.

[7:57] Vou fazer um For aqui, I variando de um até o nosso K, que é dois, e aqui dentro eu faço o quê? Chi\_2 mais igual, eu já mostro o que significa isso daqui.

[8:18] Dentro desse cara, eu vou fazer o quê? Eu vou copiar esse carinha, é isso que eu quero, é a parcela da soma. Vou dar dois pulos e vou visualizar esse resultado aqui.

[8:37] Ou seja, novamente 5,12. Isso aqui debaixo é a mesma coisa daqui de cima. Esse mais igual é exatamente a mesma coisa que eu fazer isso daqui, Chi\_2 mais, eu estou somando dentro do próprio valor.

[8:50] Estou somando e atribuindo a soma ao valor, entendeu? Aqui eu começo com zero, então zero mais isso aqui, coloco lá.

[8:57] Depois ele já está com o valor embutido, então o próximo valor mais a próxima soma eu coloco lá dentro. Eu esqueci de uma coisa aqui.

[9:05] Deu certo, não sei por que, mas a gente tem que mudar o zero para o I. E, agora sim, ele vai dar certo.

[9:16] Foi uma sorte aí a gente conseguir o mesmo valor, mas agora está certo.

[9:22] Frequência observada I, I, I. Ele vai fazer de novo para o zero, vai fazer para o um, vai atribuindo a soma e vai juntar tudo no Qui e vai dar 5,12, novamente.

[9:32] E a gente já tem aqui onde esse cara se posiciona, ou seja, na área de rejeição. 5,12 está aqui, o 3,84, a área de rejeição.

[9:44] Aqui aquela tabelinha com as situações que podem ocorrer. No caso, aqui no Qui-quadrado eu só tenho esse teste desse tipo, não tenho bicaudal.

[9:54] As frequências aqui, as hipóteses, a estatística de teste que a gente calculou e os critérios de rejeição de  $H_0$ . Como é que eu faço aqui?

[10:03] Rejeito  $H_0$  se Qui-quadrado for maior que o Qui-quadrado que eu calculei. Quer dizer, se a estatística de teste for maior que o meu Qui-quadrado Alfa.

[10:12] Está aqui, o Qui-quadrado é esse aqui, a estatística de teste que eu calculei agora aqui em cima, que é o 5,12, ela é maior que o Qui-quadrado Alfa, aquele 3,84?

[10:22] Sim, é verdade. Se isso aqui for verdadeiro, eu rejeito o  $H_0$ . E a conclusão já está aqui.

[10:26] "Com um nível de confiança de 95% rejeitamos a hipótese nula e concluímos que as frequências observadas e esperadas são discrepantes", são diferentes.

[10:39] "Ou seja, a moeda não é honesta e precisa ser substituída", senão vai atrapalhar o jogo todo. Beleza? Era isso que eu queria mostrar.

[10:46] No próximo vídeo, eu vou mostrar aquela mesma coisa do P valor, só que aqui eu também vou calcular porque o Qui-quadrado é um pouco diferente, e depois eu vou mostrar uma forma de fazer isso via Python.

[10:57] Toda essa conta que a gente fez lá em cima não é necessária, a gente faz com uma linha de código. Até o próximo vídeo.