**04 - Teste Qui-quadrado**

[0:00] Já começamos a trabalhar com testes não paramétricos e conhecemos a distribuição Qui-quadrado, que utilizaremos para resolver o nosso problema.

[0:07] Então vamos ler o problema e resolvê-lo utilizando o teste 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". 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.

|  | **CARA** | **COROA** |
| --- | --- | --- |
| Observado | 17 | 33 |
| Esperado | 25 | 25 |

[0:54] A soma disso tudo dá 50. E o esperado, obviamente, é de que a moeda, se for equilibrada, 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] Lembre-se de que no Qui-quadrado, a hipótese nula que é testada é de não haver diferença entre as frequências observadas de determinado evento, ou seja, aquelas que a gente testou, e as frequências realmente esperadas. O que eu estou esperando é que a frequência do número de caras seja igual a frequência do número de coroas.

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

F\_Observada = [17, 33]

F\_Esperada = [25, 25]

significancia = 0.05

confianca = 1 - significancia

k = 2 # Número de eventos possíveis

graus\_de\_liberdade = k - 1

COPIAR CÓDIGO

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

[1:56] Embaixo, F\_esperada, 25, 25, meio a meio para cada um. Significância: 0,05, 5%. Confiança: 1 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. Número de eventos possíveis são dois, cara ou coroa.

[2:29] Grau de liberdade era o n menos 1, agora vai ser o K menos 1, ou seja, o grau de liberdade a gente tem 1. Rodamos isso aqui, vamos para 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 nem vamos fazer o passo dois.

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

[2:56] E 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 imoport chi, ou "Qui", chame como quiser, é Qui-quadrado.

[3:23] Eu estou chamando parte da tabela que a gente criou no vídeo anterior. Só a parte que me interessa, para não ficar uma tabela muito grande, e a gente ver como consultar a tabela.

| **p** | **0.005** | **0.010** | **0.025** | **0.050** | **0.100** | **0.250** | **0.500** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Graus de Liberdade |  |  |  |  |  |  |  |
| 1 | 0.0000 | 0.0002 | 0.0010 | 0.0039 | 0.0158 | 0.1015 | 0.4549 |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.1026 | 0.2107 | 0.5754 | 1.3863 |
| 3 | 0.0717 | 0.1148 | 0.2158 | 0.3518 | 0.5844 | 1.2125 | 2.3660 |
| p | 0.750 | 0.900 | 0.975 | 0.950 | 0.990 | 0.995 |  |
| Graus de Liberdade |  |  |  |  |  |  |  |
| 1 | 1.3233 | 2.7055 | 5.0239 | 3.8415 | 6.6349 | 7.8794 |  |
| 2 | 2.7726 | 4.6052 | 7.3778 | 5.9915 | 9.2103 | 10.5966 |  |
| 3 | 4.1083 | 6.2514 | 9.3484 | 7.8147 | 11.3449 | 12.8382 |  |

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

[3:44] Lembrando que é a área de aceitação de H0. Nessa tabela, eu vou escolher um valor que vai separar a distribuição em duas partes.

[3:58] A parte da esquerda (95%) é justamente o que está mostrando aqui em cima, seria o p. Então, eu quero que tenha 95% do lado esquerdo e 5% do lado direito. Respectivamente, significância e confiança.

[4:12] Com 95% de confiança eu tenho, para um grau de liberdade, duas possibilidades, 2 menos 1 é igual a 1, portanto, eu tenho o valor de 3,8415.

[4:22] Perfeito, a gente já temos o nosso Qui-quadrado que corta o gráfico com o valor de 3,8415. Como é que eu acho isso se eu não tiver uma tabela disponível?

[4:30] Vou chamar de Qui-quadrado alfa, isto é, chi\_2\_alpha. Repare que é o mesmo procedimento que a gente vem fazendo no paramétrico, a gente vai fazer uma coisa bem parecida.

[4:41] Alfa vai ser igual a chi, chamo aquela função PPF. É a mesma coisa que a gente fez na normal, no t de Student.

chi\_2\_alpha = chi.ppf(confianca, graus\_de\_liberdade) \*\* 2

COPIAR CÓDIGO

[4:51] Eu passei confiança, vírgula, grau de liberdade. Só que aqui tem uma diferença. Lembra que a gente está trabalhando com Qui-quadrado?

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

[5:11] Lembre-se que estamos fazendo manualmente, na força, no braço, depois eu vou mostrar para você uma forma simples e rápida, usando uma linha de comando do Python.

chi\_2\_alpha = chi.ppf(confianca, graus\_de\_liberdade) \*\* 2

chi\_2\_alpha

3.8414588206941245

COPIAR CÓDIGO

[5:22] O resultado é 3,84, como a gente tinha descoberto mais acima. 3,8415, que é a linha pontilhada que divide as áreas no gráfico.

[5:29] Estatística de teste é o próximo passo. Para ela, temos uma fómula com 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 o somatório, vai de 1 a K.

[5:49] Eu coloquei i=1 aqui, porque em estatística a gente começa sempre a contar de 1,2,3 até n. Em computação, a gente vai ter que fazer um menos 1 disso, porque os índices começam com zero e vão até n - 1.

[6:02] É só a gente trocar. Isso é só uma notação estatística, uma fórmula, onde eu tenho em cada parcela a frequência observada, menos a frequência esperada, tudo isso elevado ao quadrado, dividido pela frequência esperada.

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

[6:24] Vou chamar de chi\_2, Qui-quadrado. São duas parcelas de uma soma, a primeira e a segunda.

chi\_2 = () + ()

COPIAR CÓDIGO

[6:33] Eu vou já abrir parênteses aqui dentro da primeira, porque eu tenho um numerador e um denominador.

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

[6:52] O meu primeiro índice é 1, mas lembra que em computação a gente começa com zero, então 1 quer dizer zero. Menos F\_esperada.

[7:04] Mesma coisa, o primeiro índice é zero. Isso tudo elevado ao quadrado e dividido por F\_esperado.

chi\_2 = ( (F\_Observada[0] - F\_Esperada[0]) \*\* 2 / F\_Esperada[0] ) + ()

COPIAR CÓDIGO

[7:21] São duas parcelas, como eu disse, a primeira é essa. A segunda é o outro índice. Vamos copiar essa primeira parcela toda e vamos colar dentro do segundo parêntese.

[7:35] Só precisamos trocar os índices, agora não é mais zero, é 1, o próximo valor que estamos somando. Essa é a conta que eu tenho que fazer.

chi\_2 = ( (F\_Observada[0] - F\_Esperada[0]) \*\* 2 / F\_Esperada[0] ) + ( (F\_Observada[1] - F\_Esperada[1]) \*\* 2 / F\_Esperada[1] )

chi\_2

5,12

COPIAR CÓDIGO

[7:46] O resultado é 5,12. Vou fazer de uma forma mais elegante, um pouco mais simples, usando chi\_2, eu vou atribuir um zero, porque eu estou iniciando essa variável, chi\_2 = 0.

[7:57] Vou fazer um for com i variando de 1 até o nosso K, que é dois, e dentro eu faço o chi\_2 mais igual e eu já mostro o que significa isso.

chi\_2 = 0

for i in range(k):

chi\_2 +=

COPIAR CÓDIGO

[8:18] Vou copiar a parcela da soma, depois, dar espaço de duas linhas e visualizar o resultado.

chi\_2 = 0

for i in range(k):

chi\_2 += (F\_Observada[i] - F\_Esperada[i]) \*\* 2 / F\_Esperada[i]

chi\_2

5.12

COPIAR CÓDIGO

[8:37] Ou seja, novamente 5,12. Essa fórmula chega no mesmo resultado da de cima. O mais igual é exatamente a mesma coisa que fazer chi\_2 +, eu estou somando dentro do próprio valor.

[8:50] Estou somando e atribuindo a soma ao valor. Começo com zero, então zero mais a parcela.

[8:57] Depois ele já está com o valor embutido, então o próximo valor mais a próxima soma eu coloco dentro.

[9:22] Frequência observada i, i, i. Ele vai fazer de novo para o zero, vai fazer para o 1, vai atribuindo a soma e juntar tudo no Qui e vai dar 5,12, novamente.

[9:32] Já temos onde o X² se posiciona, na área de rejeição. Ou seja, 5,12 está após o 3,84, na área de rejeição.

[9:44] Temos também uma tabela apresentando as situações que podem ocorrer. No caso, no Qui-quadrado eu só tenho um teste desse tipo, não tenho bicaudal.

[9:54] As frequências, as hipóteses, a estatística de teste que a gente calculou e os critérios de rejeição de H0. Como eu faço?

[10:03] Rejeito H0 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,chi\_2 > chi\_2\_alpha. A estatística de teste que eu calculei mais acima, que é o 5,12, é maior que o Qui-quadrado alfa, aquele 3,84?

[10:22] Sim, é verdade. Se isso for verdadeiro, eu rejeito o H0. E a conclusão já está aqui: "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. Era isso que eu queria mostrar.

[10:46] No próximo vídeo, eu vou mostrar aquela mesma coisa do P valor, mas 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 fizemos não é necessária, ela pode ser resumida a uma linha de código. Até o próximo vídeo.