**02 - Distribuição Qui-quadrado**

[0:00] Estamos começando uma seção nova do nosso curso de estatística e vamos discutir testes não paramétricos. Mas o que é isso, testes não paramétricos?

[0:08] Quando a distribuição dos meus dados amostrais não é uma normal, ou então eu não tenho elementos suficientes para afirmar que seja. E geralmente isso acontece com amostras pequenas, eu não consigo validar o teorema do limite central, aquele que a gente viu no curso anterior, fizemos até uma prova dele.

[0:24] E, nesse caso, eu não consigo utilizar métodos paramétricos para testar as minhas hipóteses, então eu tenho que utilizar testes não paramétricos, porque não temos as suposições sobre a distribuição dos dados amostrais, então eu consigo utilizar testes não paramétricos para testar as minhas hipóteses.

[0:43] Começando, como sempre, com um problema, para a gente tentar entender o porquê de utilizar esse tipo de teste que vamos conhecer, que é o Qui-quadrado. Vamos lá.

[0:52] "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, ou seja, que não tendam para determinado resultado".

[1:06] Aquela moedinha no começo do jogo que ele joga para o cara escolher o lado do jogo e a bola.

[1:12] "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."

[1:26] "A tabela abaixo mostra o resultado obtido no experimento." Aqui embaixo a gente tem uma tabelinha onde tem a cara e a coroa, que é onde a moeda caiu.

|  | **CARA** | **COROA** |
| --- | --- | --- |
| Observado | 30 | 20 |
| Esperado | 25 | 25 |

[1:32] O observado é o quê? Foram os 50 lançamentos feitos para testar, isso é o que foi observado. E o esperado é o que a gente espera.

[1:41] Se aquela moeda for legal, se ela foi muito bem organizada, ela vai ser não-viciada, ela vai cair praticamente 25 vezes para um lado e 25 vezes para o outro. Se ela for honesta.

[1:53] Estou olhando aqui a tabela, já parece que essa moeda é meio torta, então a gente vai testar justamente isso.

[1:58] Continuando o problema: "**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?"

[2:12] Testes desse tipo, a gente consegue solucionar utilizando o teste do Qui-quadrado. Que também é conhecido aqui, já deixei uns textinhos para você ler, como teste de adequação ao ajustamento.

[2:23] E ele utiliza uma variável estatística padronizada elevada ao quadrado, que é justamente essa representada por essa letra grega, o Qui, que é esse X mais sofisticado, elevado ao quadrado.

[2:35] Esse teste, está escrito aqui embaixo também para você não perder, ele testa a hipótese nula de que não existe diferença entre as frequências observadas e as esperadas.

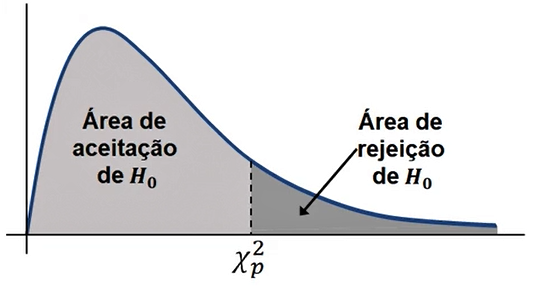
[2:48] Justamente isso que está na nossa tabela. O que a gente vai testar, isto é, a nossa hipótese nula é que não existe diferença entre essas duas frequências, do observado e do esperado.

[2:58] Para isso, a gente vai ter que utilizar os valores tabulados dessa variável, Qui-quadrado, dessa estatística padronizada do mesmo jeito que a gente fez com a normal, do mesmo jeito que a gente fez com o t de Student.

[3:14] Eu vou mostrar para vocês como calcular essa tabelinha. Ela também está atrás de todos os livros de estatística. Vou ensinar como usar essa aqui.

[3:22] Eu estou mostrando aqui para você agora, talvez no seu livro você tenha que entender como ela funciona, então preste atenção como eu estou utilizando.

[3:33] Se for igual ao seu livro, ótimo, mas se não for, você tem que entender o que a tabela está mostrando.



[3:39] A forma da função é uma coisa meio assimétrica mesmo. Temos a área de aceitação de H0, já vamos começar com isso.

[3:48] A estatística, o Qui-quadrado, o valor que a gente vai obter, do mesmo jeito que fizemos com o t Student, é a mesma ideia.

[3:55] E também temos a área de rejeição de H0. A forma de construir a tabela vou deixar aqui de presente também.

[4:01] Está tudo importado, o Pandas, o scipy.stats import chi, que é justamente essa função para obter a probabilidade.

import pandas as pd

from scipy.stats import chi

tabela\_t\_chi\_2 = pd.DataFrame(

[],

index=[i for i in range(1, 31)],

columns = [0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.975, 0.95, 0.99, 0.995]

)

for index in tabela\_t\_chi\_2.index:

for column in tabela\_t\_chi\_2.columns:

tabela\_t\_chi\_2.loc[index, column] = "{0:0.4f}".format(chi.ppf(float(column), index)\*\*2)

tabela\_t\_chi\_2.index.name='Graus de Liberdade'

tabela\_t\_chi\_2.rename\_axis(['p'], axis=1, inplace = True)

tabela\_t\_chi\_2

COPIAR CÓDIGO

[4:12] Nós temos todo esse esquema, então, não vou ficar falando disso, você pode ir analisando.

[4:17] O chi.ppf, é aquele PPF que a gente já conhece, igual ao da normal, igual ao do t student.

[4:23] Dê uma analisada para você entender como está funcionando. Rodando isso com Shift + Enter, ele vai mostrar uma tabela.

| **p** | **0.005** | **0.010** | **0.025** | **0.050** | **0.100** | **0.250** |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Graus de Liberdade |  |  |  |  |  |  |  |
| 1 | 0.0000 | 0.0002 | 0.0010 | 0.0039 | 0.0158 | 0.1015 |  |
| 2 | 0.0100 | 0.0201 | 0.0506 | 0.1026 | 0.2107 | 0.5754 |  |
| 3 | 0.0717 | 0.1148 | 0.2158 | 0.3518 | 0.5844 | 1.2125 |  |
| 4 | 0.2070 | 0.2971 | 0.4844 | 0.7107 | 1.0636 | 1.9226 |  |
| 5 | 0.4117 | 0.5543 | 0.8312 | 1.1455 | 1.6103 | 2.6746 |  |
| 6 | 0.6757 | 0.8721 | 1.2373 | 1.6354 | 2.2041 | 3.4546 |  |
| 7 | 0.9893 | 1.2390 | 1.6899 | 2.1673 | 2.8331 | 4.2549 |  |
| 8 | 1.3444 | 1.6465 | 2.1797 | 2.7326 | 3.4895 | 5.0706 |  |
| 9 | 1.7349 | 2.0879 | 2.7004 | 3.3251 | 4.1682 | 5.8988 |  |
| 10 | 2.1559 | 2.5582 | 3.2470 | 3.9403 | 4.8652 | 6.7372 |  |
| 11 | 2.6032 | 3.0535 | 3.8157 | 4.5748 | 5.5778 | 7.5841 |  |
| 12 | 3.0738 | 3.5706 | 4.4038 | 5.2260 | 6.3038 | 8.4384 |  |
| 13 | 3.5650 | 4.1069 | 5.0088 | 5.8919 | 7.0415 | 9.2991 |  |
| 14 | 4.0747 | 4.6604 | 5.6287 | 6.5706 | 7.7895 | 10.1653 |  |
| 15 | 4.6009 | 5.2293 | 6.2621 | 7.2609 | 8.5468 | 11.0365 |  |
| 16 | 5.1422 | 5.8122 | 6.9077 | 7.9616 | 9.3122 | 11.9122 |  |
| 17 | 5.6972 | 6.4078 | 7.5642 | 8.6718 | 10.0852 | 12.7919 |  |
| 18 | 6.2648 | 7.0149 | 8.2307 | 9.3905 | 10.8649 | 13.6753 |  |
| 19 | 6.8440 | 7.6327 | 8.9065 | 10.1170 | 11.6509 | 14.5620 |  |
| 20 | 7.4338 | 8.2604 | 9.5908 | 10.8508 | 12.4426 | 15.4518 |  |
| 21 | 8.0337 | 8.8972 | 10.2829 | 11.5913 | 13.2396 | 16.3444 |  |
| 22 | 8.6427 | 9.5425 | 10.9823 | 12.3380 | 14.0415 | 17.2396 |  |
| 23 | 9.2604 | 10.1957 | 11.6886 | 13.0905 | 14.8480 | 18.1373 |  |
| 25 | 46.9279 |  |  |  |  |  |  |
| 26 | 48.2899 |  |  |  |  |  |  |
| 27 | 49.6449 |  |  |  |  |  |  |
| 28 | 50.9934 |  |  |  |  |  |  |
| 29 | 52.3356 |  |  |  |  |  |  |
| 30 | 53.6720 |  |  |  |  |  |  |

[4:31] No próximo vídeo eu vou mostrar como obter isso aqui usando o PPF, seria como se eu tivesse obtendo um ponto na tabela, então não se preocupe de entender, isso é só para construir a tabela para você.

[4:44] Aqui é, basicamente, a mesma coisa da t de Student. Temos graus de liberdade, n menos 1 é o grau de liberdade, até esqueci de colocar aqui. Também temos o p, que é a área à esquerda do do Qui-quadrado e que encontraremos a partir dele e que é um dos pontos que estão na tabela.

[5:07] Por exemplo, se eu estou assumindo um p de 95%, ou seja, um nível de confiança de 95%, com um grau de liberdade 1, que é o que a gente vai utilizar aqui no nosso problema, eu vou ter um Qui-quadrado de 3,8415. Ou seja, esse valor de X²p é 3,84 e alguns quebrados.

[5:34] Ele divide isso de forma que 95% fica para um lado e os 5% ficam para o outro lado, que é o um menos p. Perfeito?

[5:42] Pessoal, a tabela é isso, a forma de consultar é essa. A gente vai estudar isso no nosso próximo vídeo, quando estivermos resolvendo esse nosso problema, utilizando o teste de Qui-quadrado. Eu vou ensinar para você como funciona, na prática, e depois como calcular, de forma simples, utilizando o Scipy, beleza? No vídeo que vem a gente resolve o problema. Abraço.