05

Distribuição binomial

Transcrição

[0:00] Legal, pessoal, vamos agora realmente iniciar nosso curso de estatística parte 2, vamos falar de distribuições de probabilidade, ou seja, distribuições teóricas de probabilidade.

[0:08] Quando a gente avalia a forma como a variável aleatória se distribui, a gente consegue definir diferentes tipos de distribuições de frequência ou probabilidade.

[0:19] Distribuições de frequência é aquilo que a gente viu no curso de estatística parte 1, você deve estar lembrado, o que vamos fazer agora, vou mostrar para vocês três distribuições teóricas de probabilidade importantes que são muito utilizadas em estatística.

[0:34] vou mostrar para vocês a distribuição binomial, a Poisson e a Normal.

[0:36] Eu vou começar lendo um probleminha, depois vamos tentar encaixar esses probleminhas nas distribuições de probabilidade para tentar solucionar eles.

[0:48] Em um concurso para preencher uma vaga de cientista de dados, temos um total de 80 questoes de múltipla escolha, cada uma com três alternativas possíveis, essas questões têm o mesmo valor, supõe que um candidato resolva se aventurar sem ter estudado absolutamente nada, e resolva fazer a prova totalmente no chute.

[1:15] Assumindo que a prova vale 10 pontos, e que a nota de corte é 5, ou seja, se o cara tirar 5 ou mais, ele passa, se tirar menos que isso, é reprovado, eu pergunto para você, qual a chance desse candidato passar para próxima etapa do processo seletivo?

[1:31] Problema desse tipo, a gente consegue resolver utilizando a distribuição de probabilidade binomial, que vamos começar a estudar a partir de agora.

[1:41] Deixei aqui um pequeno texto com explicação dessa distribuição, a gente tem a fórmula, temos o que representa cada item da fórmula para você usar com material de consulta, um experimento binomial é caracterizado pela possibilidade de ocorrência de apenas duas categorias, ou é sim ou é não, ou é verdadeiro e falso, sucesso ou fracasso, ou o cara é maior que 1 metro e 70, ou é menor que 1 metro e 70, coisas desse tipo.

[2:11] A soma dessas duas categorias é o espaço amostral, e o que é o espaço amostral? O total de eventos possíveis dentro de um experimento.

[2:22] Exemplo prático, lançamento de uma moeda é um experimento, quais são os eventos possíveis desse experimento? Ou vai dar cara, ou vai dar coroa, ok? São dois, esse é o espaço amostral.

[2:32] Outro exemplo, lançamento de um daod para cima, o dado tem seis faces, o que pode acontecer? Só pode ter seis resultados possíveis, correto? 1, 2, 3, 4, 5, ou 6, esse é o espaço amostral desse experimento, lançamento de um dado.

[2:46] Aqui, vamos à fórmula, probabilidade de ocorrer K, K é o que? É um número de eventos desejados que tenham sucesso.

[2:56] É o que estamos estudando.

[2:59] Esse NK é a combinação de NKK, logo abaixo vou mostrar como resolver esse carinha, caso alguém nunca tenha ouvido falar disso, é um método matemático bem simples de resolver, temos um método no Python que faz a continha fácil para gente.

[3:17] Passando para frente, temos P elevado a K, P é o que? Probabilidade de sucesso, de ocorrer o evento que estou estudando, por exemplo, se eu estiver estudando o número de caras que ocorrem no lançamento de uma moeda, eu quero saber a probabilidade de ocorrer cara, que é justamente o contrário, no caso que eu dei de exemplo, é a probabilidade de ocorrer coroa, ou seja, como p+q é igual a espaço amostral, ou seja, p+q é 100%, é igual a 1, ok? Em probabilidade, os valores variam de 0 até 1.

[3:51] Como P+q=1, Q é igual a 1-p, solução algébrica bem simples.

[4:04] N é o número de eventos estudados, no nosso exemplo, temos 80 questões, e dizemos que a nota de corte é 5, temos que raciocinar que se 80 questões, a nota de corte é 5, todas tem o mesmo valor, ou seja, o cara tem que acertar pelo menos 40 questões.

[4:21] N é 80, o K é o 40, o número de eventos que desejo que tenha sucesso pro cara ir para frente, no mínimo.

[4:31] Então, vamos lá, o experimento binomial tem etapas que o caracterizam, se você seguir essas etapas, consegue identificar se o problema pode ser resolvido utilizando distribuição binomial.

[4:46] Primeiro, realização de N ensaios idênticos, se quero jogar a moeda, são ensaios idênticos, a questão de três escolhas são ensaios idênticos, 3 opções cada uma delas, os ensaios são independentes, eu jogo a moeda uma vez, depois jogo uma segunda, essa segunda vez que joguei independe de eu ter jogado antes ou do que vou jogar depois, são independentes.

[5:10] Somente dois resultados são possíveis como falamos agora pouco.

[5:15] A probabilidade de sucesso é representado por P, e a de fracasso por U, de ensaio para ensaio.

[5:25] Quando jogo uma moeda, a chance de cair cara ou coroa é 50%, 50% de chance de dar um ou outro, no próximo lançamento continua sendo 50%.

[5:33] Aqui, mais uma coisa teórica, média da distribuição binomial é representado por n-p.

[5:45] Depois, essa média, vou mostrar aplicação prática disso, o desvio padrão também está aqui, sigma é igual a raiz quadrada de n vezes q.

[5:59] Agora, voltando aqui à esse NK, essa combinação de NKK, vou mostrar para vocês, primeiro, vou importar, tenho aqui, como sempre deixo, a referência para ajuda, você clicando aqui, você vai na ajuda, na documentação do Scypy, onde está essa funcionalidade que vamos utilizar agora, então vou importar de primeiro, começando com from scipy.special, ok, e aí sim, import comb.

[6:37] Esse Comb é justamente o cara que resolve o problema absurdo para gente, parece uma coisa horrorosa, mas é bem simples de resolver, a combinação de N, K, K.

[6:49] Num exemplo prático disso, por exemplo, tenho 4 amigos, e quero combinar esses amigos em pares, de 2 em 2, quantas combinações consigo fazer com esses meus 4 amigos, duas a duas? Eu posso pegar o primeiro com o segundo, depois o primeiro com o terceiro, o primeiro com o quarto, o segundo com o terceiro, o segundo com o quarto, e assim por diante.

[7:15] O que esse cara faz é te dar a resposta, quantas combinações consigo fazer com esses caras, em pares.

[7:20] Já usei um exemplo prático disso, está aqui a forma da combinação, N fatorial sobre K fatorial, vezes N-K fatorial, fatorial, o que é? É o produtório de uma contagem regressiva.

[7:38] Se eu tenho 5 fatorial, o resultado da operação é 5 vezes 4 vezes 3 vezes 2 vezes 1.

[7:47] É o produto de uma contagem regressiva, bem simples.

[7:51] Por definição, 0 fatorial é igual a 1, maravilha? Exemplo prático, bem simples, a mega-sena, todo mundo deve ter ouvido falar de mega-sena, uma loteria que tem 60 números, vamos descobrir a probabilidade de ganhar na mega-sena.

[8:13] Vamos lá, Num volante de loteria Mega-sena, temos um total de 60 números, de 1 a 60, para escolher onde a aposta mínima é 6, existem outras categorias de aposta e você paga mais, mas a mínima é 6, não me lembro o valor que você paga.

[8:29] Você que é curiosa ou curioso resolve calcular a probabilidade de se acertar na mega-sena com apenas um joguinho, você faz um volante daquele e joga, qual a chance que você tem de ganhar? Para isso, o que precisamos? Descobrir o espaço amostral, as possibilidades possíveis que temos nesse experimento da Mega-sena, e quais são? Aí justamente entra a combinação.

[9:00] O que eu preciso saber para colocar, por exemplo, a minha chance, tenho uma chance só, como diz no problema, de ganhar na mega-sena, e preciso do meu espaço amostral para fazer 1 dividido pelo espaço amostral, se eu tenho a chance de ganhar na mega-sena.

[9:15] O espaço amostral é justamente a combinação de 60 números, 6 a 6, como eu dei o exemplo de 4, 2 a 2, agora não, agora temos uma coisa maior, uma combinação de 60, seis a seis, então vamos ver qual é esse denominador da probabilidade, esse espaço amostral.

[9:37] Usando o Comb, vou chamar de combinações, combinações, é igual a comb, que acabei de importar aqui em cima, e eu só passo para ele justamente isso, desse jeito aqui, 60,6, eu quero ver quais são as combinações, repito aqui embaixo e rodo.

[10:05] Olha que absurdo, tenho 50 milhões, 63 mil, 860 resultados possíveis, combinações possíveis.

[10:15] São jogos possíveis que posso fazer, mais de 50 mil, a probabilidade de ganhar justamente eu jogando um bilhetinho, é 1 dividido pelas combinações, pelos 50 milhões, ok? Vamos fazer mais bonitinho aqui.

[10:34] Probabilidade É igual, copiar esse aqui porque senão vai ser um número que não vamos entender, vou dar um print, é um número muito pequeno, então vamos acabar perdendo informação.

[10:53] vou botar aqui 0.15, quero que tenha 15 casas decimais, e probabilidade.

[11:02] Então, vamos rodar aqui, está lá, essa é a chance que eu tenho de ganhar na mega-sena, bem pequena, é 0,0000000199 não sei o que.

[11:14] Voltando só àquela formula que vimos da combinação aqui em cima, a combinação de 60, seis a seis, é representada dessa forma aqui, 60, seis a seis, aqui 60 fatorial, dividido por 6 fatorial, multiplicado por 60 menos 6 fatorial.

[11:29] Resolvendo essa quizomba, temos esse resultado aqui, que no caso aqui do experimento da mega-sena é meu espaço amostral, quero saber minha chance de ganhar jogando um único bilhetinho, está aqui a probabilidade de eu ganhar na mega-sena é essa.

[11:48] Próximo vídeo, vamos pegar o problema que começamos e resolver utilizando a distribuição binomial, ok? Vejo você lá, abraço.