

Intervalo de confiança

Transcrição

[0:00] Legal, pessoal, vamos colocar a mão na massa agora, calcular intervalos de confiança, aprender a fazer isso, quando eu digo que estou calculando intervalo de confiança com nível de confiança que vimos na aula anterior, de 95%, eu estou querendo dizer que o parâmetro que estou estimando tem uma probabilidade de 95% de estar dentro daquele intervalo, aqui consideramos aquele erro inferencial, ou seja, isso aqui é uma estimativa não pontual mas intervalar, ok? A estimativa pontual é justamente a média, você tira uma amostra, tira a média daquela amostra, aquilo é uma estimativa pontual, o que estou colocando aqui é um intervalo onde considero os erros e dou uma probabilidade, esse erro é calculado segundo nível de confiança.

[0:52] Lembra disso? Vamos lá, aqui as fórmulas para o intervalo de confiança para a média populacional, a primeira com desvio padrão populacional conhecido, isso pode ocorrer, e a segunda com desvio padrão não conhecido, vamos fazer usando o desvio padrão da própria amostra, perfeito? Aqui, no primeiro caso, μ representa parâmetro populacional, a média populacional, igual a \bar{x} , que é o que, eu faço a seleção da amostra, tiro a média da amostra, esse carinha é o \bar{x} , perfeito? Mais ou menos, por que o mais ou menos? Vou ter um intervalo inferior e superior, vou calcular dois números, para falar a verdade, o primeiro, intervalo inferior, e o segundo, intervalo superior, vamos ver como isso funciona.

[1:39] Mais ou menos, aquele erro inferencial que obtemos na última aula, Ezinho, Z , a raiz de sigma sobre a raiz de N , essa fórmula, por isso que vemos vendo o teorema do limite central, aprendemos o que isso significa, já tínhamos visto probabilidade, aprendemos a obter o Z , vamos calcular de novo para reforçar.

[1:59] Aquele probleminha que eu joguei lá no começo, quando a gente começou a falar de estimação, suponha que os pesos do saco de arroz de indústria alimentícia se distribuem aproximadamente como uma normal de desvio padrão populacional, está passando para mim, desvio padrão populacional, igual a 150 gramas, selecionada uma amostra aleatória de 20 sacos, de um lote específico, obteve-se peso médio de 5 mil e 50 gramas.

[2:30] Construa o intervalo de confiança para a média populacional com nível de significância de 5%, a gente aprendeu no ultimo vídeo, nível de significância e confiança, quando digo significância de 5%, o que quero dizer? Que o nível de confiança é de 95%, 1 menos alfa, esse é alfa, então vamos lá, coloquei aqui, preparei informações que o problema deu para a gente, deu uma média amostral, 5 mil e 50.

[2:56] Aqui em cima, média amostral, o nível de significância, igual a 0.5, e vou calcular meu nível de confiança, prefiro trabalhar com nível de confiança, 1 - alfa, ou seja, confiança, igual a 1 menos significância, rodei, 95, perfeito, deixei preparado aqui a tabela normal padronizada, a que temos lá em cima, mas fiz seleção nela, repare, 16:26, quero algumas linhas porque vou mostrar para vocês uma coisinha, e vou passar para vocês uma tabelinha com níveis de confiança mais utilizados, você pode consultar essa tabelinha também, logo aqui embaixo, o que a gente precisa, olhando a formula, obter? Precisamos do Z , o sigma, e o raiz de N , perfeito?

[3:51] Lógico, \bar{x} foi dado, como obtenho o Z ? Deixei aqui uma foto com o Z pronto, 1.96, mas como obtenho esse Z ? Lembra que lá quando estávamos trabalhando com a distribuição normal, lembra que tínhamos o Z aqui, mas o Z era calculado, tinha abaixo dele, toda essa área hachurada, lembra disso? Ou seja, daqui do meio, para cá, a gente sabe que vale 0.5, legal? Daqui para cá, tinha que calcular, como eu tenho uma função espelhada, simétrica, eu sei que esse lado aqui do meio até aqui, quando tenho 95%, tenho que fazer o que? Pegar esse 95, 0.95, e dividir por 2, eu acho aquela areazinha ali, concorda? Ou seja, aqui no meio até aqui, temos 47.5% ou 0.475, a mesma coisa, o que está faltando aqui?

[5:05] Lá, na minha tabela da normal, aqui, eu preciso saber qual é esse valor todo, porque a tabela é representada desse ponto para cá, perfeito? Como que faço isso para pegar daqui para cá, e vou obter essa probabilidade, essa área hachurada,

que não é mais 95%, vou mostrar para você, é daqui até o final, contando com esse vazio, obtendo essa área toda, desse ponto até aqui, vou na minha tabela e consulto esse valor aqui, como faço isso? Como obtive essa área, sei que aqui vale 0.475, o que falta para obter o resto todo? Justamente 0.5, metade, ok? Vamos fazer essa conta, 0.5, mais essa conta aqui de cima, perfeito? É assim que a gente calcula essa probabilidade ali, 0.975, ou seja, a área sob a curva normal, deste ponto que, vamos supor que eu não saiba que é 1.96, deste ponto até o final, que é menos infinito, é 0.975.

[6:31] Eu vou lá, na tabela, e vou procurar 0.975, vamos lá, a tabela me diz que 0.975, vou procurar aqui dentro, 9744, 0.975, perfeito? Eu preciso do Z para esse valor, como faço? Esse cara aqui, na linha, mais esse da coluna, lembra disso? 1.9 mais 0.06.

[6:58] Vou fazer aqui, 1.9 mais 0.06, ou seja, 1.96, justamente esse cara que tem aqui.

[7:11] Como é função simétrica, espelhada, esse valor vai ser exatamente igual a esse, com sinal invertido, perfeito? Já concluímos esse tópico.

[7:25] Para obter o Z de outra forma, sem ser via tabela, temos também a ajuda do nosso amigo Python, do Scipy, você lembra, importamos o scipy.stats, Z vai ser igual a 9, mas aqui vai dar aproximação, uma coisa mais sofisticada, mais exata, ok? Norm, eu chamo Norm, tínhamos obtido lá utilizando CDF, no CDF passo o Z e ele me dá a probabilidade, eu passo esse Z e ele me dá isso aqui, aqui eu quero o contrário, eu passo isso aqui, e que ele me dê o Z, o método que usamos é PPF, aquela ajuda que deixei para vocês arrumadinho, passo aqui dentro para eles, essa probabilidade aqui, vamos supor que eu não tenha tabela, passo para ele a probabilidade.

[8:26] Desculpa, não mostrei, 1.9599, ou seja, aproximadamente 1.96, aqui ele é mais exato, perfeito? Aqui embaixo, deixei para vocês uma tabela justamente com os mais comuns, 90, 95 e 99%, níveis de confiança, aqui, nível de confiança, aqui o valor da área sob a curva normal, o que estávamos calculando, 95, 97.5 para 95 e 99.5 para 99% de nível de confiança, e os valores já arredondados, caso você queira usar, mais rápido, mais prático, aqui a tabelinha com os mais comuns.

[9:07] Voltando, vamos calcular, obter o sigma x barra, que é justamente o sigma sobre a raiz de N, vamos calcular ele, qual o desvio padrão? Desvio padrão no nosso problema, de uma amostra, ele deu que era aqui, desvio padrão populacional, 150, passou para gente essa informação, então desvio padrão populacional igual a 150, perfeito? Vamos só manter o padrão de nosso código aqui, desvio padrão, 150.

[9:46] N, a gente sabe que é 20, está bom? Também está lá, deixa eu rodar aqui antes, essa amostrinha aqui, amostra aleatória com 20 sacos, N = 20, aqui embaixo, estávamos aqui. O que preciso fazer? Vamos calcular aqui a raiz de N, é o que, NP, a gente chama o numpy, sqrt é a função para calcular uma raiz quadrada, e passo o N, perfeito? Raiz de N, pronto, estamos com todos os elementos para calcularmos esse sigma x barra aqui em cima.

[10:36] Que é o que, desvio padrão, vou chamar de sigma, apenas sigma, desculpem, igual a desvio padrão dividido por esse cara, raiz de N, tá? Pronto, temos mais sigma, mais um elemento da nossa fórmula, o que está faltando? Obter o erro, o que chamo de erro? Essa parte da fórmula aqui, esse cara aqui, Z vezes sigma sobre raiz de N, é o errinho aqui, ok? Vamos calcular, temos os elementos para calcular, bem simples, vamos calcular o E, o E vai ser igual a Z, perfeito, vezes sigma, perfeito?

[11:34] Está tudo pronto, o E é esse carinha aqui, 65.73, que unidade isso está mesmo? Gramas, né? Gramas, estamos falando de sacos, grãos, 65 gramas.

[11:50] É isso que a gente quer, vou mostrar duas formas de calcular intervalo de confiança, uma manual, já que vemos fazendo todo o processo manual, depois mostro uma forma simples onde você passa alguns parâmetros e tem um output.

[12:04] Primeiro, vamos chamar de intervalo, essa é a forma manual, eu vou abrir e fechar, mas vou dar um pulo aqui, vamos chamar, lembra da média, copiar aqui em cima, só para ir relembrando o que já fizemos, isso é dado do problema, média

amostral, ele passou para gente, calculou para gente média amostral, justamente o \bar{x} barra da nossa fórmula, vamos lá, média amostral, primeiro intervalo, lembra do mais ou menos, limite inferior, vamos utilizar menos, menos o que? Menos o erro, justamente esse cara aqui em cima, é bom a gente sempre ir e voltar, a gente tem \bar{x} barra, mais ou menos o erro que calculamos, então, como temos o limite inferior primeiro, vamos calcular o menos, quando não é vírgula aqui, faz a mesma coisa embaixo, alterando por sinal, invertendo ele, menos e mais.

[13:11] Deixa eu botar aqui para ter uma visualização de intervalo, ok, está aí, nosso intervalo de confiança é esse, com nível de confiança de 95%, ou seja, eu tenho uma probabilidade de 95% de o parâmetro da minha população estar dentro do intervalo.

[13:35] A forma automática de calcular isso, lembra do Norm, Norm.interval.

[13:45] Eu passo para ele o alfa, na verdade 1 menos alfa, 0.95, eu passo para ele o loc, um parâmetro que representa o \bar{x} barra, a média, que é 5 mil e 50, mas vamos botar para você saber do que estamos falando, media, e o scale, que é justamente o sigma.

[14:17] Esse carinho que a gente obteve, ou seja, estou passando para ele o que? A nossa fórmula, nossos dados da fórmula, \bar{X} barra, calcula o Z internamente e o errinho para ele que é nosso sigma aqui, está bom? Esse carinho eu passo para ele também, vamos lá, voltando aqui, o sigma, perfeito, pode ser assim sem problema.

[14:49] Reparem, meus resultados aqui em cima, perfeito? Essa é outra forma de obtermos o intervalo de confiança de forma mais automática, perfeito? Era isso que eu queria mostrar para vocês, pessoal, no próximo vídeo, vamos entrar em outra sessão e vamos, utilizando informações que tivemos nesses vídeos na parte de estimação, vamos aprender a calcular, obter tamanho de amostra para que a amostra seja realmente representativa da população, perfeito? É isso que a gente vai ver nos próximos vídeos, até lá, abraço.