

Cálculo para a população infinita

Transcrição

[0:00] Pessoal, maravilha, vamos começar uma sessão nova, vamos aprender a calcular o tamanho da amostra, para que minha amostra seja realmente representativa da população que estamos estudando, a gente já aprendeu a estudar erro inferencial e vamos começar a utilizar isso a partir de agora também, aprendemos a obter o Z, distribuição de probabilidade normal, a partir de uma probabilidade, conseguimos obter o Z, lembra disso? Também vamos utilizar essa ferramenta nas formas de cálculo de tamanho de amostra, perfeito? Vou começar com o problema clássico que sempre começamos as sessões, esse é o seguinte, estamos estudando rendimento mensal dos chefes de domicílio no Brasil.

[0:43] Estamos falando do dataset que abrimos no começo do nosso curso.

[0:49] Nosso supervisor determinou que o erro máximo em relação a média seja de 100 reais, erro máximo, aqui está dando o erro máximo de 100 reais.

[0:59] Sabemos que o desvio padrão populacional desse grupo de trabalhadores é de 3 mil, 323 reais, e 39 centavos, para um nível de confiança de 95%, qual deve ser o tamanho da amostra de nosso estudo? Eu trouxe lá de cima, novamente, a fórmula do erro inferencial, porque a fórmula para calcular o tamanho de uma amostra é derivada dessa fórmula, tenho duas fórmulas diferentes, porque uma é para o desvio padrão conhecido, usam o sigma, e uma pro desvio padrão desconhecido, eu uso estimativa dele, o S, ok, a gente viu isso falando do intervalo de confiança, as duas formas distintas, justamente por causa disso.

[1:44] Reparem que trouxe a fórmula do erro inferencial, porque como eu disse, essa fórmula é derivada da outra, a única coisa que fizemos foi isolar o N algebricamente, a gente passou aqui para o outro lado da igualdade, descemos o E aqui pra baixo, repare aqui, raiz de N igual a uma outra expressão, a gente, pra eliminar a raiz do N, eleva os dois lados da equação ao quadrado, no caso do N a gente elimina a raiz e aqui fica ao quadrado, ou seja, N igual a Z vezes sigma sobre E, que é o erro, elevado tudo ao quadrado.

[2:24] A diferença da outra fórmula é que no lugar de sigma a gente usa o S, que é estimativa do sigma, perfeito? É isso, dois pontos que devem ser, que sempre devemos prestar atenção quando formos aplicar esse tipo de fórmula, desvio padrão, tanto sigma quanto S, eles tem que estar na mesma unidade de medida do erro, deixei observações pra você lembrar disso, desvio padrão e erro devem estar na mesma unidade de medidas.

[2:53] Outra coisa, quando o erro, a gente tiver expressão do erro em percentual, esse erro é um percentual em cima da média, também está aqui embaixo, representado em termos percentuais, deve ser interpretado com percentual relacionado a média, perfeito? Com isso na cabeça, conseguimos aplicar a fórmula tranquilamente.

[3:19] Voltamos ao exemplo que acabamos de ler, estudando rendimento mensal dos chefes de domicílio no Brasil, o supervisor determinou que o erro máximo em relação à média seja de 100 reais, e temos um desvio padrão populacional de 3 mil, 323 reais, e 39 centavos, nível de confiança de 95%, qual o tamanho da amostra? Vamos começar a calcular isso, primeira coisa, vou obter o Z, simplesmente aplicação de fórmula, esse cara aqui, vou usar esse cara aqui de cima, ele me deu o desvio padrão populacional, tenho que usar sigma, vamos lá.

[3:56] Obter o Z, vamos começar por isso, Z vai ser igual, já obtemos lá em cima, norm.ppf, a mesma coisa que fizemos lá, 95%, vamos repetir a conta aqui em cima pra não ficar nenhuma dúvida, voltar rapidamente, só pra gente relembrar do que estou falando, esse cálculo que vamos fazer aqui embaixo de novo, tá vendo? Tenho que passar esse 95 para minha função norm.ppf, ok? Vamos relembrar, rapidamente, coisa rápida, o que tenho, 0.95, como função é simétrica, quero só uma parte dela, divido por 2, perfeito? Tem aqui 0.475.

[4:54] Só que eu quero todo aquela área, quero daquele ponto, até o menos infinito, o que tenho que fazer? Somar 0.5, então, 0.5 mais esse cara aqui de cima, já fizemos isso, só para reforçar, e vamos ter o valor de 0.975, assim que calculamos o outro, fizemos novamente, passamos a informação como argumento, para norm.ppf, e aqui a gente obtém o Z de forma fácil, 1.9599, ou seja, 1.96.

[5:33] Perfeito, Sigma, mais uma informação da fórmula, sigma vai ser igual, isso é um dado do problema, aqui, 3 mil, 323, e 39.

[5:46] Ok, continuando, vamos rodar isso para manter na memória, mantendo padrão, 3323.39.

[6:00] O erro também, pelo que lembro, foi dado no problema, lembra que o cara falou, o supervisor, no máximo 100 reais em cima da média, aqui, erro máximo em relação à média, seja de 100 reais, ele deu um intervalo, criou intervalo de confiança para a gente, ok? Tem que colocar tudo ali dentro, e rodamos o erro, o que temos que fazer agora? Aplicar fórmula, só isso.

[6:27] Qual a fórmula? N igual a Z vezes sigma sobre e elevado ao quadrado, é isso, perfeito.

[6:35] N é igual a Z, a gente atribuiu o valor a Z, vezes, aqui, vou fazer um parêntese para ficar bem divididinho, sigma dividido por e, perfeito? Tudo isso aqui, asterisco, asterisco, exponencial, elevado ao quadrado, N é igual a 4 mil, vamos fazer o seguinte, para não ficar número quebrado, estamos falando de indivíduos, vamos arredondar isso, n.round.

[7:18] Perfeito, Pronto, a gente arredonda pra cima, sem menor problema, 4 mil, 243 é o número de elementos que preciso ter na amostra para garantir tudo que foi pedido no problema, nível de confiança, de 95%, um erro máximo em relação à média de 100 reais, estou garantindo tudo isso com esse tamanho de amostra, ela é representativa da população, garantindo todos esses pontos, considerando desvio padrão populacional, informação importante também, pessoal, é isso aqui, lembra, aplicação dessa fórmula, variáveis quantitativas e população infinita, identificou esse tipo de característica no estudo, corre pra essa fórmula, no próximo vídeo, vamos ver outra forma, já falamos de variáveis também quantitativas mas para população finita, próximo vídeo, abraço.