

02

Praticando: Graus de Liberdade

Olá!

Na lousa, eu mostrei pra vocês a diferença de população e amostra. E mostrei que amostra é o que nós geralmente temos em mãos quando vamos fazer um estudo. E mostrei também que a fórmula da variância, em particular, muda se for pra amostra ou pra população. A gente coloca um $n-1$ no divisor. E a gente subtrai 1 por quê? Porque nós acreditamos que a variância da população seja maior do que a da amostra. E aí esse -1 faz com que o número fique maior, e a gente acredita que isso é mais próximo da variância da população. Esse $n-1$ é o que nós chamamos, em inglês, de **degrees of freedom (graus de liberdade)**. E essa aula em particular eu não vou mostrar no R, até porque não tem muita coisa pra mostrar, mas vou mostrar pra vocês de onde veio esse $n-1$.

Então, graus de liberdade, degrees of freedom, eu falei pra vocês que é $n-1$. Imagine o seguinte exemplo: se eu disser pra você que eu tenho quatro números que eu não sei quais são: x, y, z e t (planilha do excel, respectivamente: A2, B2, C2, D2). Só que eu sei que a soma de $x+y+z+t = 10$. Aqui o excel não faz, obviamente, porque é variável. Mas, eu sei que a soma tem que ser 10. A minha pergunta é, que número eu posso pôr em x ? Se são quatro números, e x é o primeiro número. Que número eu posso pôr em x ? Qualquer um. O número que eu escolher não tem problema, se eu escolher 600, eu vou poder ter uma combinação de outros três números cuja soma vai me dar 10. Vamos supor que o primeiro seja 3 ($x=3$). O próximo número agora, que número pode ser? Também qualquer um. O número que eu escolher, eu vou poder dar um jeito de que os próximos dois corrijam pra dar 10. Vamos supor que o próximo seja 2 ($y=2$). Agora, eu estou no terceiro número. Que número pode ser? Também qualquer um. Vamos supor que seja 1 ($z=1$). Ótimo. Agora, o quarto número, que número pode ser? Aqui, veja que eu não tenho escolha. Se eu sei que a soma é 10, pra esse número aqui, dado que os anteriores são 3, 2 e 1, esse aqui tem que ser 4. Pra que a soma de $A2+B2+C2+D2$ seja 10. Veja que, de quatro números, eu tinha três que podia escolher à vontade: isso são os graus de liberdade, degrees of freedom. E você vê que isso aparece em vários momentos da estatística.

Na fórmula da variância, em particular, vamos entender o que é que acontece com esses graus de liberdade. Porque é isso que a gente está fazendo, dando um grau de liberdade que é o $n-1$. A gente tem um grau a menos de liberdade do que a gente usa lá na fórmula da população. Imagine que a soma eu já saiba, porque ali na variância a fórmula é a soma da diferença do número pra média ao quadrado, dividido por n . Então, imagine que eu já saiba que a soma daquilo é 240. E vamos imaginar que $n=4$. Se $n=4$, na fórmula da população vai ficar $240/4$. E na fórmula da amostra, vai ficar $240/3$. Dá uma olhada em como ficou: da população é 60, da amostra é 80.

$n=4$ População: $240/4 = 60$ Amostra: $240/3 = 80$

Vamos aumentar esse número para $n=8$.

*População: $240/8 = 30$

*Amostra: $240/7 = 34,28571429$

Se aumentar mais ainda, $n=20$

*População: $240/20 = 12$

*Amostra: $240/19 = 12,6317895$

Veja só, antes era 60 e 80, a diferença era grande; 30 e 34. Quando foi pra $n=20$, ficou 12 e 12,6 já está bem perto. Agora, dá uma olhada em $n=120$

n=120

*População: 240/120 = 2

*Amostra: 240/119 = 2,016806723

A diferença já caiu pra 01. O que isso quer dizer pra gente? Quer dizer que se a sua amostra tem mais de 120 pessoas eu nem preciso muito mais olhar aquele "-1", porque eu já estou vendo que os números são parecidos. E por quê eu estou dizendo isso? Eu estou dizendo que dependendo do número de pessoas que você tem no seu estudo, de elementos que você colheu ali, você não precisa nem se preocupar em usar o n-1 na fórmula. Isso porque os dois números serão muito parecidos. Então era isso que eu queria mostrar pra vocês aqui: que tem aquela fórmula lá da amostra, em que a gente faz o n-1, só que ela faz sentido quando a sua amostra é muito pequena. Quando a sua amostra é grande, os números serão parecidos tanto da fórmula que você usa de maneira normal, pra variância, quanto da fórmula que você usa pra amostra. E esse n-1 é o que nós chamamos de graus de liberdade.

E agora você entendeu também, embora eu não vá entrar no detalhe matemático da coisa, de quando usamos o tempo inteiro... Mas quando alguém diz que tem 4 graus de liberdade, quer dizer que 4 variáveis estão livres, podem ser mudadas à vontade. Foi exemplo que eu dei pra vocês com o numeral 10 no início. Pensa, nesse nosso último exemplo, que aquele n-1 é o grau de liberdade pra chegarmos nessa soma de 240. Se eu tenho n-1 graus de liberdade, quer dizer que eu tenho n-1 números que eu posso mudar, e só o último tem que ser fixo ali.

É isso que eu queria mostrar na aula, então não fique muito preocupado se você usar a fórmula da variância comum, dependendo do tamanho da sua amostra, porque vai dar mais ou menos na mesma, tá legal?

Obrigado!

