

誤差率について

佐藤健一（広島大・原医研）ksatoh@hiroshima-u.ac.jp

誤差率は正式には標準誤差率とよばれ、標準誤差を平均で割った値として定義されます。標準誤差が平均の値に対してどの程度大きいものかを見えています。同様の概念としては標準偏差を平均で割った変動係数などがあります。

皆さんの疑問は、今のデータで十分な精度が確保できているか？だと思いますが、次のある町で行われた無作為抽出による成人男性の身長調査を例として考えてみましょう（データは架空のものです）。

統計量	略記号または式	値
標本数	n	688
平均	μ	152.82
標準偏差		17.57
標準誤差	$SE = \frac{\text{標準偏差}}{\sqrt{n}}$	0.670
標準誤差率	$100 \times (SE \div \mu)$	0.438
95%信頼区間	$(\mu - 1.96SE, \mu + 1.96SE)$	(151.51, 154.13)

誤差率を眺める前に見慣れている平均の 95%信頼区間を見えます。95%信頼区間の意味は「神様しか知らない身長の真の平均（母平均）がこの区間に入っている確率は 95%である」となります。もう少し詳しく説明すると、『この町の成人男性全員の身長を測れば真の平均は分かるが、あまりにお金がかかるから、無作為に選んだ n 人について身長を測った。もちろん、もう一度同じように集めればきっと平均もかわるだろう。それでも真の平均のありかについて、ある程度は知ることができる』となります。入ってない確率も 5%はあるわけですが、どうやら、真の身長は 151 センチから 154 センチ程度と言えそうです。

同じ平均 152.82 が得られたとしても標本数によって信頼区間は伸び縮みします。もし、標本数が少なければ平均の 95%信頼区間が(100, 200)ということもあり得ます。あまりに範囲が広すぎてなんの目安にもなりません。逆に、標本数が増えれば、平均の 95%信頼区間が(151.51, 154.13) となるかも知れません。少数 1 桁目まで、身長の真の平均を絞り込むなんてすごい精度です。

このようにして、標本数を増やせば真の平均のありかを絞り込むことができます。今のデータでは 95%信頼区間は(151.51, 154.13) です。真の平均は 151, 152, 153, 154 あたりには在りそうです。この精度をどう思いますか？十分ですか？大雑把すぎますか？もし、これが許容できるなら、貴方にとって十分な精度の調査ということになります。

話は誤差率に戻ります。いつも同じような集団で調査を行っている方なら、誤差率の目標値（例えば 0.5%など）があるかも知れません。ふつうは、同時に計測された他の項目との精度比較に利用されるようです。例えば、「体重も身長と同時に調査されたけど、何人かに拒否されたために標本数もかわり、その結果、誤差率が 1.2 となった」とすると、体重の推定精度は身長よりも悪い、と判断されるようです。標準誤差率は概念的には信頼区間と同じですが、無単位なので単位系が異なる項目について推定精度が比較できるメリットがあると言えそうです。