

百均の電卓で解ける 統計入門

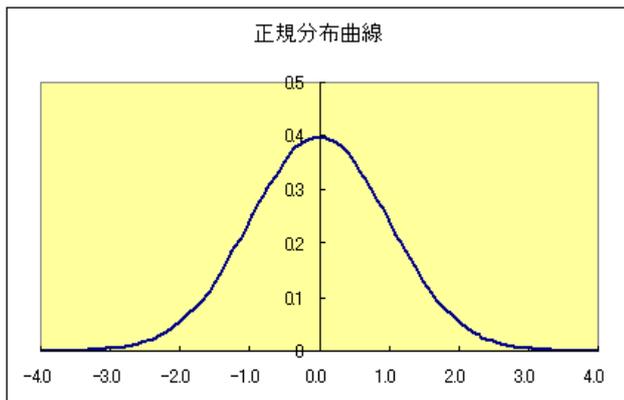
数学的な厳密さよりも直感的に理解することに
重点を置いていきます・・・

正規分布は予言の書？

統計的推定？

前回、平均値を中心として ± 1 S.D. の値をとるのは全体の 68.26% であり、 ± 2 S.D. には 95.44%、 ± 3 S.D. には 99.74% のものが含まれることをお話ししました。・・・と、言う事は、ある測定データが正規分布に従う時、あらかじめ出るデータを予言出来る？

例えば、子供の年齢がわかれば、2、3年後のその子の身長を予言・・・。また、数年後の結果を見て、「想定内の身長だよー。」なんて言うこともあるでしょう。この予言が『統計的推定』の入口なのです。



正規分布曲線は、上図でしたね。このとき、山の高さは、データの現れる相対度数を表していたのですから、最も山の高い 0.0 の値を予言すれば、最も高確率に当たるということです。

しかし、どうでしょう、「2年後の身長は〇〇cm」と1つの値を指定して当たる確率・・・。

確かに山のピークは他よりも確率は高いとは言え、まだまだ低いですよ。

そこで、予言にも幅を持たせて見ましょう。「2年後の身長は〇〇cm～〇〇cm」といった具合です。この時、図の ± 1.0 に該当する値を言えば、68.26% 当たる確立になりますよね。

95%的中？

さて、どの程度予言が当たれば良いのでしょうか？

確かに、100% 当たれば良いのですが、「2年後のこの子の身長は 0cm～250cm」なんて予言、当たるでしょうけど・・・。

一般的に統計を扱う場合、「95%的中」あるいは「99%的中」を多用します。ここでは、「95%的中」で説明していきます。

図の ± 2.0 の範囲には、95.44% が入ることは説明してきましたので、大体の予想はつくと思いますが、「95%的中」する範囲は、

$$-1.96 \text{ 以上 } +1.96 \text{ 以下}$$

となります。

でも、ちょっと待ってください。「95%的中」する範囲なら他にもありますよね。例えば、「-2.1 以上 +1.86 以下」もその一例でしょう。

この時、何が違うのでしょうか？

予言する幅を考えてみてください。前者は 3.92、後者は 3.96 となります。予言するのであれば、範囲は小さい方が良いわけで、前者を使うのです。最も短い範囲は、0.0 を中心に左右対称の選択になります。

平均値と S.D. がわかったら？

いままで、図のような標準正規分布で説明してきましたが、一般正規分布ではどうなるのでしょうか。

標準正規分布は、平均値 = 0、S.D. = 1 であったことを思い出してください。一般正規分布は、標準正規分布のすべてのデータに一定数を掛け、その後一定数を足すことで求められます。

掛ける一定数を σ (ギリシャ文字でシグマ)、加える一定数を μ (ギリシャ文字でミュー) として、

(一般正規分布のデータ) =

$$\sigma \times (\text{標準正規分布のデータ}) + \mu$$

となりましたね。

なので、平均値が μ で S.D. が σ の一般正規分布の「95%的中」する範囲は、

$$(\mu - 1.96\sigma) \text{ 以上 } (\mu + 1.96\sigma) \text{ 以下}$$

となります。

なので、一般的に「95%的中」するためには「 ± 1.96 S.D. 以内の範囲を予言すればいい。」こととなります。

ただし、この時に「5%はハズレる。」ことを忘れないでください。

この式を使うと、コイン投げの予言で表が 10 枚出ることが 95% 当たるのは、15 枚投げた時？ 20 枚？ 25 枚？ 30 枚？ 35 枚？ なんてことも出来るのですが・・・。

【町田幸雄】 この項続く



シクラメン

不老長寿・勇敢



ポインセチア

私は燃えている



寒椿

謙讓、愛嬌

12月の花

