

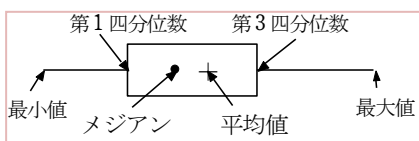
雑感 箱ひげ図の能力

■ データの様子を見るとき、ヒストグラムは最(?)強の表現手段であると思われる。しかし、ヒストグラムは「図」であるだけに、扱いづらいという弱点がある。そのため、その集団の状態を手取り早く、数値で見るための手段として、「代表値」があるのだと思っている。

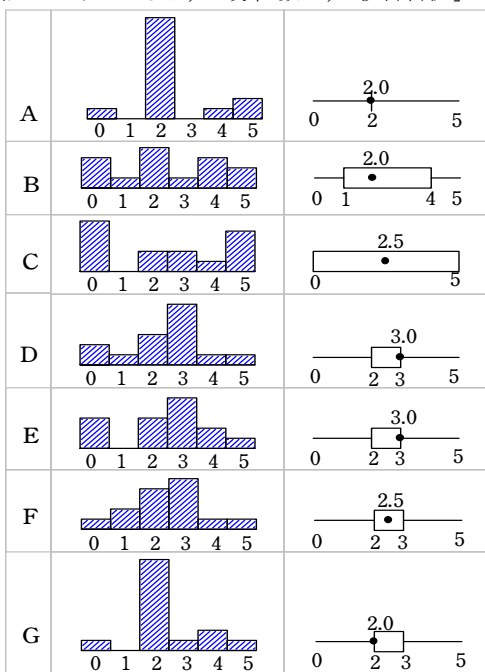
■ 代表値の代表が「平均値」であり、他にメジアン(中央値)やモード(最頻値)も用いられることがある。

さらに、標準偏差、レンジ(範囲)や、数学Iの教科書に登場することになった四分位数も調べると、集団の様子をよく詳しく見るができるということらしい。

■ 箱ひげ図は、一般的に次のように代表値などを記入した図式であり、ヒストグラムの代わりとなるものという位置づけであろう。



■ さて、箱ひげ図はどれだけ集団の状態を表現できるのだろうか。試みに、データ数14で、平均値を2.4、最小値を0、最大値を5に揃えたA~Gの集団を作ってみた。そして、それらのヒストグラムと箱ひげ図(平均値2.4は同じなので記入を省いた)を併記してみた。なお、四分位数は、「教科書流」に依った。



■ A, B, Cの違いは、箱ひげ図でもその違いとして表現されている。一方、D, Eは箱ひげ図が完全に一致する集団であるが、ヒストグラムもそれなりに似ていると言っていいただろうか。

それに対して、箱ひげ図がそれなりに似ているE~Gについて、ヒストグラムはずいぶん違っているように思われる。メジアンの相違がここまで関係しているとは、驚きである。

こうなると、分散の値が特徴を明らかにするために意味があるのだろうか。ちなみに、各集団の分散の値は次の通りである。

A : 1.8, B : 3.2, C : 4.6, D : 2.0, E : 2.4, F : 1.6, G : 1.5

う〜ん、FとGは分散も値が近いなあ。

■ 修行を積めば、箱ひげ図からヒストグラムを予想するようになるのだろうか。それとも、そもそも箱ひげ図の能力は、精々この程度のものなのだろうか。