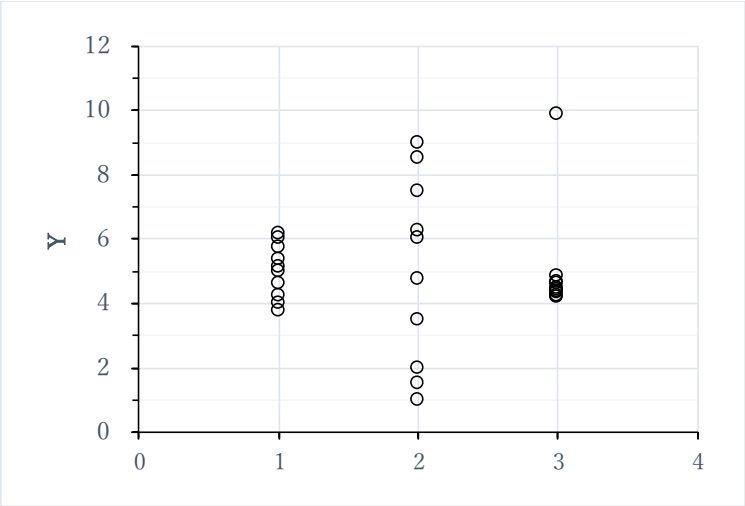
#002★☆☆☆ 2020.3.3 / 2023.7.21

平均値や標準偏差の数字だけをみないで、個々のデータのグラフからばらつきを確認すべき。

いわゆる「要約統計量」といわれる平均値や標準偏差は論文などにもよく示されており、データがおおよそどのような値になるのかを知る目安になる。しかし要約された数値だけ見ていると間違った判断をしてしまうこともあり、ミスを避けるためには必ず生データのプロットを眺めるべきである。

下表のデータを EXCEL で入力し、右図のような個々のデータのプロットを作成してみる。グラフは便宜上、横軸を A、B、C ではなく 1、2、3 としている。

	A	B	C
1	4.00	1.00	4.25
2	3.75	1.50	4.36
3	4.25	3.50	4.42
4	4.60	2.00	4.60
5	6.00	9.00	4.85
6	5.75	6.25	4.17
7	6.15	7.50	4.33
8	5.15	6.00	4.65
9	5.35	8.50	4.47
10	5.00	4.75	9.87
MEAN	5.00	5.00	5.00
S.D.	0.84	2.92	1.72



A、B、C どの組でも平均値は 5.0 で同じで標準偏差はそれぞれ異なっている。標準偏差の数値だけを見てしまうと、B より C のほうがデータのばらつきが小さいと判断してしまうが、図をみると C には極端に大きな数値（9.87）が含まれており、こういったデータを「外（はず）れ値」と呼ぶ。平均値や標準偏差といった数値だけみてしまうとこのような外れ値を見落としてしまうことがある。

外れ値かどうかを厳密に決定するには、もとのデータがどのような分布をしているのか（正規分布など）を考慮して、そこから統計的に見て有意に乖離しているかどうかを検定するが、実際問題としては、実験的な操作ミス（静脈血に点滴液が混じってしまった、とか）を考えて、明らかに外れ値と判断できる根拠がない限りはデータ解析に含めなければならない。根拠なく得られたデータを省いて解析してはならない。統計解析的には、外れ値がある場合、すなわちデータの分布がなんらかの仮定する統計分布にならない場合にはノンパラメトリック検定といった検定方法があり、データをみて適切な検定方法を選ぶことも大切になるが、詳しくは別の機会で触れる。