1　データの水準による選択

ピアソンの積率相関係数は、データの平均値、標準偏差、共分散の計算を伴います。従って、間隔尺度データ、比尺度データでなければなりません。

スピアマンの順位相関係数は、順序尺度のデータであっても計算できます（間隔尺度、比尺度の場合にも計算できます)。

2　直線相関と曲線相関

二つのデータの関係は、直線的なものの他に何らかの曲線で表される場合もあります。図 1 の点線で示した曲線は

$$y = \frac{100}{1 + \exp(-x + 5)}$$

で表され、一一個のデータ点はその曲線上にあります。

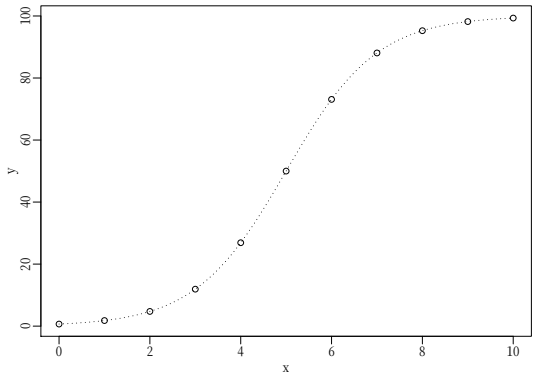

図 1　曲線関係にある二つのデータの散布図 (1)

表 1　曲線上のデータ点の座標 (1)

x	0.00	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00
y	0.67	1.80	4.74	11.92	26.89	50.00	73.11	88.08	95.26	98.20	99.33

表 1 のデータからピアソンの積率相関係数を計算してみると、〇・九七〇になります。非常に強い相関関係があるといえる結果ですが、x と y には数式で表される完全な関係があるので相関係数は一になっても良さそうに思えます。実は、ピアソンの積率相関係数は、二つのデータが $y = ax + b$ のように直線関係にあるときに相関係数が一になるのです。二つのデータ間にどのような完全な数学的関係があっても、直線関係でないかぎり二つのデータ間のピアソンの積率相関係数は一にはなりえないのです。

では、図 1 のようなデータであっても相関を表す数値が一になるようなものはないのでしょうか。あります。スピアマンの相関係数というのがそれです。スピアマンの順位相関係数というのは名前が表すとおり、順位に基づいた相関係数ということです。表 1 のデータをそのまま用いてピアソンの積率相関係数を計算するのではなく、二つのデータそれぞれについてまず小さい順に番号を振ります。そして、その番号をデータだと思ってピアソンの積率相関係数を計算するのです。表 1 のデータは、x も y も左から右へと小さい順に並んでいるので、x と y の順序番号は同じになります。つまり、「y の順序番号 $= 1 \times x$ の順序番号 $+ 0$」ですから、ピアソンの積率相関係数を計算するときのデータは直線関係です。従って、計算されるピアソンの積率相関係数（実際にはスピアマンの順位相関係数を計算しています）は一になります。

しかし、だからといってどのような関数関係にある二つのデータであってもその間のスピアマンの順位相関係数が常に一になるわけではありません。図 2 の点線で表した曲線は

$$y = 1.327x^3 - 19.813x^2 + 75.465x - 0.585$$

という三次式曲線です。一一個のデータ点の座標は表 2 のようになっています。x と y の座標と共に、それぞれの順位も表に含めています。これをみると、x は小さいに並んでいますが、y の順番は x とは異なっています。つまり x と y は直線関係ではないので、元の x と y から計算されるピアソンの積率相関係数も、順位から計算されるスピアマンの順位相関係数も一ではないことがわかります（それぞれ、〇・〇四七と〇・〇三六で、ほとんど無相関です）。

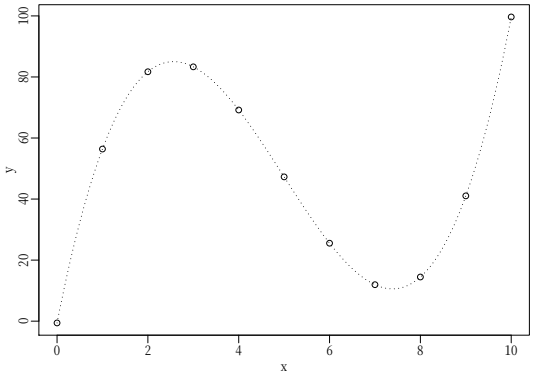

図 2　曲線関係にある二つのデータの散布図 (2)

表 2　曲線上のデータ点の座標 (2)

x	0.00	1.00	2.00	3.00	4.00	5.00	6.00	7.00	8.00	9.00	10.00
y	-0.58	56.39	81.71	83.32	69.18	47.27	25.54	11.95	14.47	41.05	99.66
y の順位	1	7	9	10	8	6	4	2	3	5	11

なお、図 2 のようなデータの場合にも相関をあらわす数値が一になるものとして、重相関係数というのがあります。この場合の重相関係数は、x の他に、x^3、x^2 という変数をひとまとめにして考えたものと y との相関を表すものです。重相関係数はピアソンの積率相関係数やスピアマンの順位相関係数と違って、〇から一の範囲の値をとります。

3　外れ値と相関係数

平均値を計算するとき、極端に大きい（あるいは小さい）データ（外れ値と呼びます）が混じっていると、それに引きずられて不適切な平均値が得られます。相関係数の場合にも同じようなことが生じます。図 3 のような場合、白丸だけのデータの相関係数はマイナス〇・八六七で強い負の相関関係があります。しかし、右上の黒丸で示したデータ一個があるだけで相関係数は〇・五六九となりかなり強い正の相関関係であるということになります。

図 4 のような場合、右上にある黒丸で示したデータ点を含めて一四個のデータから計算されるピアソンの積率相関係数は〇・九九二です。右上の黒丸が右上方向へ離れれば離れるほどピアソンの積率相関係数は一に近づいていきます。このデータに対してスピアマンの順位相関係数を計算するとわずか〇・一八五ということになります。その理由は、スピアマンの順位相関係数は、データに順位を付けてその順位にもとづいてピアソンの積率相関係数を計算するからです。右上にあった黒丸を左下の方向へ移動した図 5 を考えればよいでしょう。この二つの図において、白丸に対する順位は同じで、黒丸についても x も y も一五番目のデータであることに違いはないということです。つまり、両方の図に対するスピアマンの順位相関係数は共に〇・二二七で同

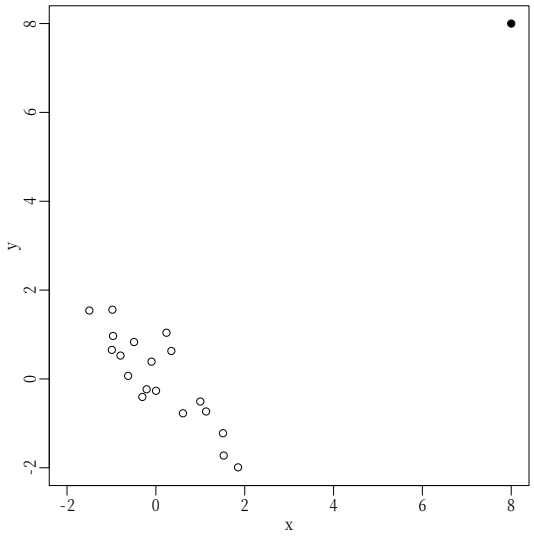

図 3　外れ値の影響

じです。ちなみに、ピアソンの積率相関係数は図 4 のデータに対しては〇・九九二ですが、図 5 においては〇・一八五となっています。

スピアマンの順位相関係数は、外れ値の影響を受けにくいという特徴があるのです。

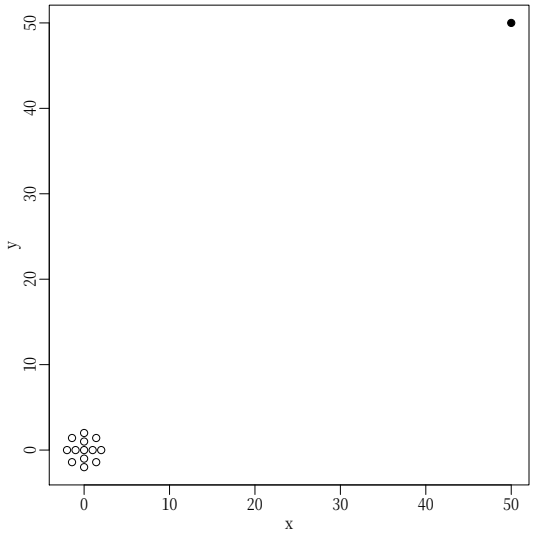

図 4　外れ値の影響

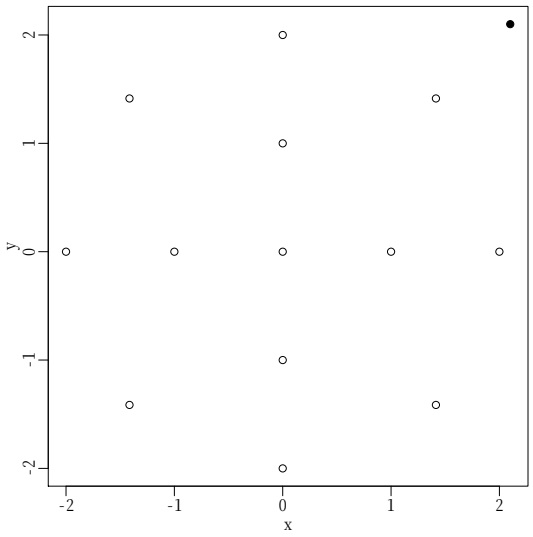

図 5　外れ値の影響を軽減する