

(5)相関係数

様々な統計的解析においては、複数の変数間の関係に着目した分析を行うことが多い。たとえば、表 1-3 のような 2 つのデータが対になっている 10 個のデータがあったとする。この場合、A に関する 10 個のデータの平均は 28.98、分散は 68.3、B に関する 10 個のデータの平均は 4.09、分散は 3.1 と計算される。

表 1-3 10 組のデータペアの例

順番	1	2	3	4	5	6	7	8	9	10	平均	分散
A	25.1	41.4	10.6	24.3	26.3	32.9	36.5	34.8	33.5	24.4	28.98	68.34
B	4.3	6.2	1.9	4.6	1.9	3.3	7.6	3.3	5.2	2.6	4.09	3.14
AB 偏差積	-0.81	26.21	40.25	-2.39	5.87	-3.10	26.40	-4.60	5.02	6.82	9.97	—

図 1-7 は、A、B 2 つの変数の対の関係を散布図に示したものである。この図を見ると、A の値が大きいたまには B の値も大きい、という傾向が見て取れる。このような 2 つの変数間の関係を定量的に示す指標として、相関係数がある。

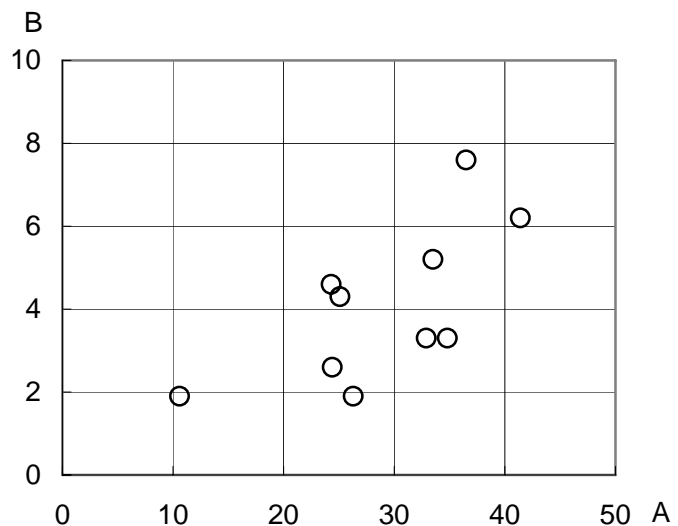


図 1-7 変数 A と変数 B の散布図

相関係数とは、(A の偏差)×(B の偏差)の平均を A の標準偏差と B の標準偏差で割ったものである。つまり、たとえば表 1-3 の順番 1 のデータは、A の値が 25.1 なので A の偏差は(A-平均)=(25.1-28.98)、B の値が 4.3 なので B の偏差は(B-平均)=(4.3-4.09)となり、(A の偏差)×(B の偏差)=(A-平均)×(B-平均)=(25.1-28.98)×(4.3-4.09)=-0.81 となる。このような計算を 10 個のデータ全てについて計算したものが「AB 偏差積」の行に示されている。これを平均すると 9.97 である。相関係数は、(AB 偏差積の平均)÷(√Aの分散 × √Bの分散)で与えられるので、A の標準偏差=√Aの分散 = √68.34 =

8.27, B の標準偏差 = \sqrt{B} の分散 = $\sqrt{3.14} = 1.77$ を用いて, 相関係数 = $9.97 \div (8.27 \times 1.77) = 0.68$ と計算される.

相関係数は $-1 \sim +1$ の範囲の値しか取ることができない. 2 つの変数に正の相関 (片方の変数が大きくなるともう一方の変数も大きくなる) があると, 相関係数の値は正の大きな値, すなわち $+1$ に近づく. 逆に 2 つの変数に逆の相関 (片方の変数が大きくなるともう一方の変数が小さくなる) があると, 相関係数の値は負の大きな値, すなわち -1 に近づく. 両者の相関が小さいほど 0 に近い値となる. 以上より, 相関係数は, 2 つの変数間の相関関係をわかりやすく定量的に示す指標である.

ただし, 相関係数は直線的な相関関係を示す指標のため, 2 変数間の直線的でない関係は必ずしも反映されない. したがって相関係数が 0 に近くても 2 変数間に関係がないとは言えない. また, 相関係数はあくまでも 2 変数間の相関関係を表すものであり因果関係を示すものではない. 従って, 変数 A, B の相関係数が 1 に近いからといって, 「A は B の原因である」とは言い切れない.