

■ 2つの変数間の関連

一方の変数の値と他方の変数の値の間に一定の傾向が見られるとき、2つの変数間に関連があると言う。

- ex) 勉強時間とテスト得点…勉強時間が長い人ほど、テスト得点が高い。
- 海藻類の摂取とイレウスの生起…海藻類を食べた患者ほどイレウスを起こす。

■ 相関係数

2つの量的変数間の直線的な関係の程度を表す指標。(曲線関係は表せない)

一方の変数の値が高いほどもう一方の変数の値が高い(低い)という傾向  
前者を正の相関、後者を負の相関、どちらの傾向もない場合、無相関と言う。

正の相関があるデータの傾向

- 一方の変数の値が平均より大きければ、他方の変数の値も平均より大きいだろう。
- 一方の変数の値が平均より小さければ、他方の変数の値も平均より小さいだろう。
- 一方の変数の平均偏差得点が+であれば、他方の変数の平均偏差得点も+だろう。
- 一方の変数の平均偏差得点が-であれば、他方の変数の平均偏差得点も-だろう。
- 2つの平均偏差得点の積は+だろう (+×+ → + -×- → +)
- 中には、一方の平均偏差得点が+のとき、他方の平均偏差得点が-になるデータもあるかもしれない。しかし、2つの平均偏差得点の積を平均すれば+になるだろう。
- さらに、直線関係が強いほど、(-のものは少なくなるし)、大きなプラスの値になるだろう。

2つの平均偏差得点の積の平均 = 2つの変数の共分散

$$\begin{aligned} \text{分散} &= 1 / \text{被験者数} \times (\text{データ} - \text{平均値})^2 \\ &= 1 / \text{被験者数} \times \begin{matrix} \uparrow \\ \text{データ} - \text{平均値} \end{matrix} \times \begin{matrix} \uparrow \\ \text{データ} - \text{平均値} \end{matrix} \\ &\hspace{10em} \uparrow \hspace{2em} \text{同じ変数} \hspace{2em} \uparrow \end{aligned}$$

$$\begin{aligned} \text{共分散} &= 1 / \text{被験者数} \times (\text{データ1} - \text{平均値1}) \times (\text{データ2} - \text{平均値2}) \\ &\hspace{10em} \uparrow \hspace{2em} \text{違う変数} \hspace{2em} \uparrow \end{aligned}$$

分散は、同一変数同士の共分散と見ることもできる。

- 正の相関があるデータにおいては、2つの変数の共分散の値は+になるだろう。
- 同様に、負の相関があるデータにおいては、2つの変数の共分散の値は-になるだろう。
- 共分散の値の符号と大きさが、正の相関、負の相関の程度の判断材料になる。

ピアソンの積率相関係数

r = 2変数の共分散 / 各変数のSDの積

-1 ~ +1 の範囲。 -1 : 完全な負の相関, 0 : 無相関, +1 : 完全な正の相関

## ■相関係数の大きさの解釈（心理・教育など社会科学系での）

- 0.0～±0.2 無相関かほとんど相関無し
- ±0.2～±0.3 相関無しか、あったとしてもわずか
- ±0.3～±0.4 弱い相関
- ±0.4～±0.6 中程度の相関 … 散布図を見て相関ありと思えるのはこの辺から
- ±0.6～±0.8 強い相関
- ±0.8～±1.0 非常に強い相関

## ■相関係数に関する注意点

- ・外れ値に弱い。極端な値のデータの混入により、相関係数の値が大きく変化してしまう。
- ・2つの変数の直線的な関係を表す。U（逆U）字相関のような曲線的な関係は表せない。  
ex) 自己開示度と好感度、開示しないのもしすぎるのも悪く、ほどよい開示がよい。
- ・1つの相関係数の値が安定してくるのは、経験的に言って、被験者数が150程度以上、300以上あっても、それほどの効果はない。
- ・相関関係は因果関係ではない。  
「一方の変数の値が高い人は他方の変数の値も高い」というのと、「一方の変数の値が高まれば、他方の変数の値も高まる」というのは異なる。  
本を良く読む親の子どもほど成績がよい、という相関関係があったとしても、親が本をたくさん読んだら子どもの成績が上がるわけではない。

## 疑似相関（見かけの相関）

対象としている2つの量的変数以外の変数の影響で、対象としている変数間の相関関係がおかしく見えること。

ex) 小学生において、足が大きいほど算数がよくできる、というほは、学年という共変数があるから。

疾患に関する知識量と合併症の進行度。同じ罹患年数の患者であれば、知識が多いほど合併症は少ないという負の相関があるが、罹患年数という共変数があるため、知識量と合併症に正の相関が見られる。

→不自然な相関（または無相関）が観察された場合、大発見をしたと思う前に疑似相関を疑うべし。→層別分析