

1. 変数の標準化
2. 相関係数
3. 相関係数行列
4. 欠損値の処理

# 【変数の標準化】

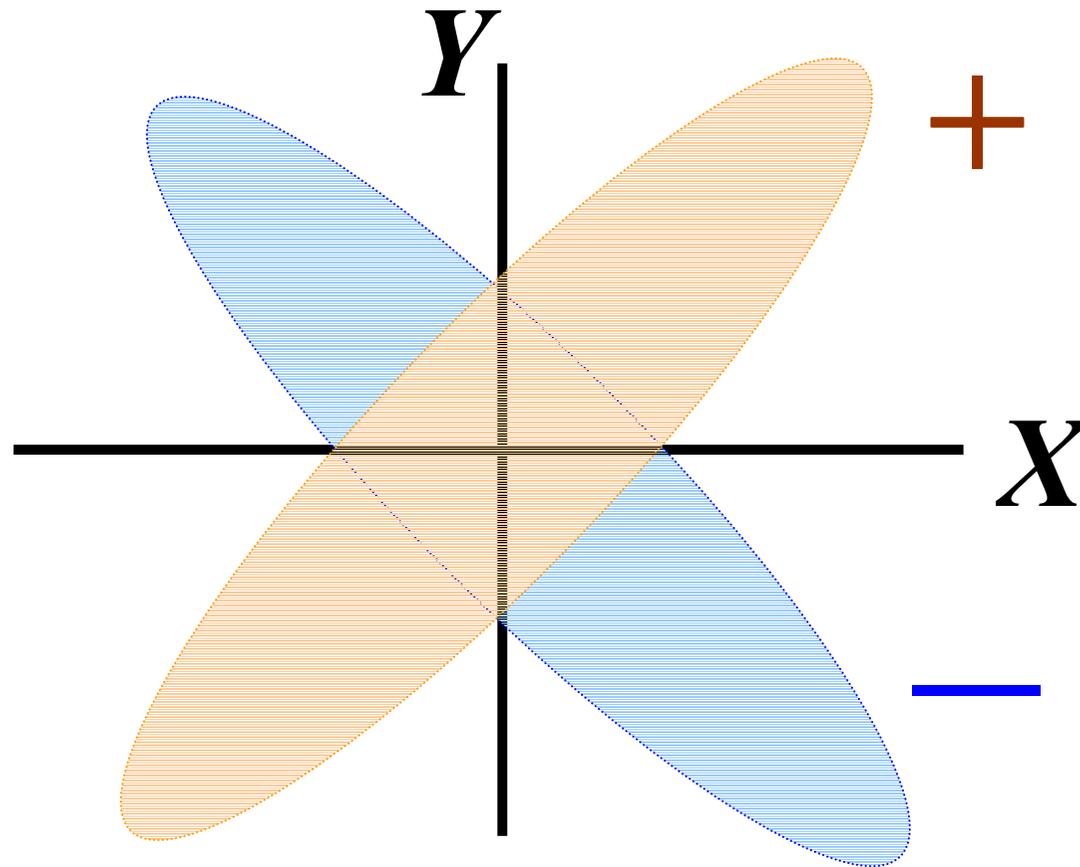
平均 = 0, 標準偏差 = 1  
になるように変換する。

$$X = \frac{x - \text{平均}}{\text{SD}}$$

これで単位を気にせずに比較できるようになる

# 【変数間の関係】

正(+ )の関係か、負(-)の関係か



# 【相関係数】

Pearson の積率相関係数

(教科書 p. 19)

(Product-moment correlation coefficient)

標準化済みの変数  $X, Y$  について

$$r = \frac{XY \text{の総和}}{N}$$

今日では、単に「相関係数」といえばこの  $r$  をさす

$r$  は  $-1 \sim +1$  の範囲の値をとる:

- ・ 無関連のときゼロ
- ・ 完全に直線的な関連のとき  $\pm 1$

解釈の目安:

- ・ 0.1 以下なら無関連、
- ・ 0.2 程度で弱い相関、
- ・ 0.4 程度で強い相関、
- ・ 0.7 をこえると「ほとんどおなじ変数」

# 【SPSS コマンド】

「相関」 → 「2 変量」

変数を指定する

## 【相関係数行列】

3 変数以上の相関係数を総当たりで出すこともできる (correlation matrix)

ただし、「オプション」で「リストごとに除去」をえらぶ

# 【欠損値の処理】

- 対単位 (pairwise) の除去  
個々の組み合わせごとに欠損ケースを除く
- 表単位 (listwise) の除去  
分析に使う変数に**ひとつでも**欠損のある  
ケースを除く

多変量解析の際は listwise 処理がふつう。

ただし、欠損ケースが多くなる。  
1割以上が欠損になるようなら注意。

## 【注意事項】

- ★ 相関係数は、はずれ値や歪みに弱い(平均値と同様)。順位相関係数 (Spearman, Kendall など) のほうがいいことも。
- ★ クロス表や散布図で関連のかたちをチェックしておくことがのぞましい
- ★ 相関係数の 95%信頼区間は  
 $N=100$  で  $\pm 0.20$ ,  $N=200$  で  $\pm 0.14$  くらい