確率・統計　後期　第11回

相関

稲積　泰宏（いなづみ　やすひろ）

今日の内容

- 相関とは
- 共分散
- 相関係数
- 相関係数の計算
- 注意点

相関とは

相関：2つの変数の間の関係の強さと方向

相関の種類

- **正の相関**：一方が増えると他方も増える傾向
  - 例：身長と体重、勉強時間と成績
- **負の相関**：一方が増えると他方は減る傾向
  - 例：気温とコート売上、運動量と体脂肪率
- **無相関**：2つの変数に関係が見られない
  - 例：身長と数学の成績

散布図と相関

散布図：2変数のデータを平面上に点で表したグラフ

- 横軸に一方の変数、縦軸にもう一方の変数をとる
- 点の並び方から相関の有無や強さを視覚的に判断

散布図のパターン

- 右上がり → 正の相関
- 右下がり → 負の相関
- バラバラ → 無相関

共分散

共分散：2つの変数の関係を数値で表す指標

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

計算公式

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n} x_i y_i - \bar{x}\bar{y}$$

共分散の性質

- $s_{xy} > 0$：正の相関（同じ方向に変動）
- $s_{xy} < 0$：負の相関（逆方向に変動）
- $s_{xy} = 0$：無相関

共分散の問題点

共分散の値の大きさは比較できない

共分散の値は、データの単位やスケールに依存する

- 身長(cm)と体重(kg)の共分散：大きな値
- 身長(m)と体重(kg)の共分散：小さな値

→ 同じデータでも単位を変えると共分散の値が変わる

解決策：共分散を標準化した指標が必要 → **相関係数**

相関係数

相関係数：共分散を標準化した指標

$$r = \frac{s_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

ここで、s_x、s_yはそれぞれx、yの標準偏差

相関係数の特徴

- 単位に依存しない
- 常に $-1 \leq r \leq 1$

相関係数の性質

相関係数の値の範囲

$$-1 \leq r \leq 1$$

相関係数の解釈

- $r = 1$：完全な正の相関（すべての点が右上がりの直線上）
- $r = -1$：完全な負の相関（すべての点が右下がりの直線上）
- $r = 0$：無相関
- $0 < r < 1$：正の相関（rが1に近いほど強い）
- $-1 < r < 0$：負の相関（rが-1に近いほど強い）

相関の強さの目安

相関係数の絶対値による解釈

- $0.0 \leqq |r| < 0.2$：ほとんど相関なし
- $0.2 \leqq |r| < 0.4$：弱い相関
- $0.4 \leqq |r| < 0.7$：中程度の相関
- $0.7 \leqq |r| < 1.0$：強い相関
- $|r| = 1.0$：完全な相関

> **注意**
>
> これはあくまで目安。分野や文脈によって解釈は異なる

例題1）

注意点：外れ値の影響

外れ値の影響

相関係数は外れ値（極端な値）の影響を受けやすい

- 1つの外れ値で相関係数が大きく変わる可能性
- データに外れ値がないか確認が重要
- 散布図で視覚的に確認するのが有効

注意点：非線形関係

相関係数は線形関係のみを測る

2変数に強い関係があっても、それが非線形なら相関係数は小さくなる

例

- $y = x^2$のような関係（放物線）
- 強い関係があるが、相関係数は必ずしも高くない

散布図で関係の形を確認することが重要

問1)

まとめ

相関

- 2つの変数の関係の強さと方向を表す
- 共分散：関係を数値化（単位に依存）
- 相関係数：共分散を標準化（$-1 \leq r \leq 1$）

重要な注意点

- 外れ値の影響を受けやすい
- 線形関係を測る

次回までに問題集Basic90-91を解いておいてください

感想を会議のチャット欄へ