

第6回 量的データの分布の記述

■ 量的データの分布の記述

分布パラメタ：量的データの分布の様子を表現する簡潔な指標の総称

分布のどのような様子を表すか

おもに、位置、散らばり（広がり）、左右対称性からの歪み、とんがり度

■ 代表値

量的データの分布の位置を簡潔に記述する値

平均値（mean）：データの重心。データ値の総和／データ数。外れ値に弱い但便利

平均偏差得点：データ - 平均値

統計学上、極めて重要なもの。現に、分散やSDの計算で使われている。

中央値（median）：データを大きさ順にならべて中央にくる値。外れ値に強い（頑健）

最頻値（mode）：出現頻度が最も多いデータ値。外れ値に強い（頑健）

単峰で、

分布が左右対称なら、3つの値は一致する → 平均値 = 中央値 = 最頻値

左に裾が長い = 値の小さい方のデータが多い → 平均値 < 中央値 < 最頻値

右に裾が長い = 値の大きい方のデータが多い → 平均値 > 中央値 > 最頻値

■ 散らばり（散布度）

量的データの分布の広がり具合を表す値

分散（variance）：（データ - 平均値）の2乗 / データ数

（データ - 平均値）の2乗の平均値

標準偏差（standard deviation, SD）： $\sqrt{\text{（データ - 平均値）の2乗} / \text{データ数}} = \sqrt{\text{分散}}$

統計ソフトなどで計算される値は不偏分散の平方根であることが多い → 推測統計

$\sqrt{\{(\text{データ} - \text{平均値}) \text{の} 2 \text{乗} / (\text{データ数} - 1)\}}$

単位の共通化

ex) データ cm, 平均 cm, 分散 cm^2 , SD cm. 平均とSDは同じ単位上の指標になっている。

統計ソフトなどで計算される値は不偏分散であることが多い → 推測統計

（データ - 平均値）の2乗 / （データ数 - 1）

平均偏差： | データ - 平均値 | の総和 / データ数

平均偏差得点の絶対値の平均

平均偏差は直感的でわかりやすいが、数学的取扱が面倒なので、ほとんど用いられない。

■分布の歪み

分布の左右対称性を表す値

歪度 (skewness) $\{(\text{データ} - \text{平均値}) / \text{SD}\}$ の 3 乗 / データ数

負に大きい : 左に裾が長い. 値の小さい方のデータが多い.

0 : 左右対称.

正に大きい : 右に裾が長い. 値の大きい方のデータが多い.

■分布のとんがり度

分布がどの程度とんがっているかを表す値

尖度 (kurtosis) $\{(\text{データ} - \text{平均値}) / \text{SD}\}$ の 4 乗 / データ数 - 3

負に大きい : 正規分布より平ら

0 : 正規分布と同じとんがり度

正に大きい : 正規分布よりもとんがっている